

No.21  
MARZO DE 2017

# Documentos CEDE

ISSN 1657-7191 Edición electrónica.

I evade taxes, and so what?  
A new database and evidence from  
Colombia

---

Leopoldo Fergusson  
Carlos Molina  
Juan Felipe Riaño

**CEDE**  
CENTRO DE ESTUDIOS SOBRE DESARROLLO ECONÓMICO

 Universidad de  
**los Andes**  
Facultad de Economía

**CEDE**  
CENTRO DE ESTUDIOS SOBRE DESARROLLO ECONÓMICO

 Universidad de  
**los Andes**  
Facultad de Economía

Serie Documentos Cede, 2017-21  
ISSN 1657-7191 Edición electrónica.  
Marzo de 2017

© 2017, Universidad de los Andes, Facultad de Economía,  
CEDE. Calle 19A No. 1 – 37 Este, Bloque W.  
Bogotá, D. C., Colombia Teléfonos: 3394949- 3394999,  
extensiones 2400, 2049, 3233  
infocede@uniandes.edu.co  
<http://economia.uniandes.edu.co>

Impreso en Colombia – Printed in Colombia

La serie de Documentos de Trabajo CEDE se circula con propósitos de discusión y divulgación. Los artículos no han sido evaluados por pares ni sujetos a ningún tipo de evaluación formal por parte del equipo de trabajo del CEDE.

El contenido de la presente publicación se encuentra protegido por las normas internacionales y nacionales vigentes sobre propiedad intelectual, por tanto su utilización, reproducción, comunicación pública, transformación, distribución, alquiler, préstamo público e importación, total o parcial, en todo o en parte, en formato impreso, digital o en cualquier formato conocido o por conocer, se encuentran prohibidos, y sólo serán lícitos en la medida en que se cuente con la autorización previa y expresa por escrito del autor o titular. Las limitaciones y excepciones al Derecho de Autor, sólo serán aplicables en la medida en que se den dentro de los denominados Usos Honrados (Fair use), estén previa y expresamente establecidas, no causen un grave e injustificado perjuicio a los intereses legítimos del autor o titular, y no atenten contra la normal explotación de la obra.

Universidad de los Andes | Vigilada Mineducación  
Reconocimiento como Universidad: Decreto 1297 del 30 de mayo de 1964. Reconocimiento personería jurídica: Resolución 28 del 23 de febrero de 1949 Minjusticia.

# I evade taxes, and so what?

## A new database and evidence from Colombia\*

Leopoldo Fergusson<sup>†</sup>      Carlos Molina<sup>‡</sup>      Juan Felipe Riaño<sup>§</sup>

### Abstract

Tax evasion lies at the core of the relationship between citizens and the state: it reflects the level of trust in the state and compliance with society's implicit 'social contract'. However, empirically analyzing it is challenging, with few direct and reliable measures. This has hampered the advancement of the theoretical and empirical literature, which is especially underdeveloped in the case of indirect tax evasion. We conduct list experiments on a large sample of households to estimate the incidence of value added tax (VAT) evasion, as well as the extent of social desirability bias in respondent answers. Around 20% of respondents engage in evasion and, surprisingly, they are not ashamed to recognize this openly. Evasion is more prevalent in places with more informality and less physical presence of the state, as well as among poorer, less educated individuals, and those who disregard the rule of law.

**Keywords:** Tax evasion, value added tax, social desirability bias, list experiments.

**JEL:** C83, C93, D73, H26.

---

\*We thank the entire team from the *Encuesta Longitudinal Colombiana de la Universidad de los Andes*, particularly its academic committee. We are especially indebted to Ximena Cadena, the project director during the design and implementation of the experiment described in this paper, for both successfully leading the survey and for numerous comments and suggestions on the survey design and our main findings. We also thank Kelley Friel for detailed comments.

<sup>†</sup>Universidad de los Andes, Department of Economics, Carrera 1ra N° 18A - 12 Bogotá, Colombia. [lfergusson@uniandes.edu.co](mailto:lfergusson@uniandes.edu.co), corresponding author.

<sup>‡</sup>Universidad de los Andes, Department of Economics, Carrera 1ra N° 18A - 12 Bogotá, Colombia. [ca.molina11@uniandes.edu.co](mailto:ca.molina11@uniandes.edu.co).

<sup>§</sup>University of British Columbia, Vancouver School of Economics, 6000 Iona Drive Vancouver, Canada. [jf.riano@alumni.ubc.ca](mailto:jf.riano@alumni.ubc.ca).

# Evado impuestos, ¿y qué?

## Una nueva base de datos y evidencia para Colombia\*

Leopoldo Fergusson<sup>†</sup>    Carlos Molina<sup>‡</sup>    Juan Felipe Riaño<sup>§</sup>

### Abstract

La evasión de impuestos está en el centro de la relación entre los ciudadanos y el Estado: refleja la confianza general en el Estado y la aceptación del ‘contrato social’ implícito en la sociedad. Sin embargo, analizar la evasión empíricamente es difícil, dada la escasez de medidas directas y confiables. Esto ha obstaculizado el avance de la literatura teórica y empírica, especialmente subdesarrollada en el caso de la evasión de impuestos indirectos. En este trabajo aplicamos experimentos de lista a una muestra amplia de hogares para estimar la incidencia de la evasión del impuesto al valor agregado (IVA) y el sesgo de deseabilidad social en las respuestas. Alrededor del 20% de los encuestados evaden y, sorprendentemente, no se sienten avergonzados de reconocerlo abiertamente. La evasión es más frecuente en los lugares con más informalidad y menos presencia física del estado, así como entre los individuos más pobres y menos educados, y aquellos que desconocen el imperio de la ley.

**Palabras clave:** Evasión de impuestos, impuesto al valor agregado, sesgo de deseabilidad social, experimentos de lista.

**JEL:** C83, C93, D73, H26.

---

\*Agradecemos al equipo de la *Encuesta Longitudinal Colombiana de la Universidad de los Andes*, especialmente a su comité académico. Estamos particularmente en deuda con Ximena Cadena, directora de la encuesta durante el diseño e implementación de los experimentos descritos en este artículo, tanto por liderar exitosamente la encuesta como por numerosos comentarios y sugerencias sobre el diseño de las preguntas y sobre nuestros resultados principales. Agradecemos también a Kelley Friel por comentarios detallados.

<sup>†</sup>Universidad de los Andes, Facultad de Economía, Carrera 1ra N° 18A - 12 Bogotá, Colombia. [lfergusson@uniandes.edu.co](mailto:lfergusson@uniandes.edu.co)

<sup>‡</sup>Universidad de los Andes, Facultad de Economía, Carrera 1ra N° 18A - 12 Bogotá, Colombia. [ca.molina11@uniandes.edu.co](mailto:ca.molina11@uniandes.edu.co)

<sup>§</sup>University of British Columbia, Vancouver School of Economics, 6000 Iona Drive Vancouver, Canada. [jf.riano@alumni.ubc.ca](mailto:jf.riano@alumni.ubc.ca)

# 1 Introduction

This paper examines tax evasion, a phenomenon that is key to understanding the nature of the state and of democratic institutions. Tax evasion not only reflects the incidence of a specific crime. Decisions about whether or not to pay taxes are also influenced by levels of general trust in the state and compliance with society’s implicit ‘social contract’: citizens pay taxes, and the state works for citizens and delivers public goods. A long tradition in the social sciences also relates state capacity to the development of an effective tax capacity.<sup>1</sup> The incidence of tax evasion thus captures the state’s enforcement ability and its capacity to mobilize resources, as well as citizens’ resistance of the state. In other words, it is related to the state’s *consensual strength* (Acemoglu, 2005). Consensually strong states do not simply have *power*, as many dictatorships might. They have *legitimate power*, because their actions respond to citizens’ needs and demands.

Despite its importance, examining tax evasion empirically is challenging. Given the nature of the behavior, measures are often noisy, too aggregate, or simply unavailable. Also, questions in surveys may fail to elicit honest answers from respondents. To address these concerns, we apply list experiments to a large sample of households to measure (value added) tax (VAT) evasion. Our data is drawn from the 2013 round of the *Encuesta Longitudinal Colombiana de la Universidad de los Andes (Elca)* (Bernal et al., 2014), the first large-scale household panel survey in Colombia, with roughly 10,000 households that are representative of urban Colombia and five rural macro regions.<sup>2</sup> A companion paper (Fergusson, Molina, & Riaño, 2017) describes a similar analysis drawing from the same source to study an equally crucial behavior in democracies: clientelistic vote buying. Also, tax evasion (or state capacity more generally) and clientelism may interact with each other, creating multiple feedback loops leading from state weakness to clientelism, and vice versa (Fergusson, Molina, & Robinson, 2017).

We pursue three main objectives. First, we estimate the prevalence of tax evasion while avoiding possible biases in citizen responses. Our list experiments protect respondents from directly admitting that they evade taxes. Instead, we randomly assign part of our sample to a treatment group that is enquired about the *number* of activities that they regularly follow to save on their expenses, out of a list containing VAT evasion as one of the options.

---

<sup>1</sup> See Besley and Persson (2009) for a discussion.

<sup>2</sup>The data are freely available for download from the project’s website, at <https://encuestalongitudinal.uniandes.edu.co/en/>. The 2013 round contains the list experiment described in this paper, and baseline characteristics are also available for 2010 panel. A new round, in 2016, collected additional information, though time and resource constraints precluded the inclusion of the list experiment. However, a direct question was included in 2016, and results will be available to the public in early 2018.

A control group is confronted to a similar prompt and list, except that evasion is not one of the alternatives. Since respondents are randomly assigned, the gap in the number of actions reported by each group estimates the proportion evading the VAT. We find that nearly one out of every five people engage regularly in VAT evasion. Crucially, to pursue our second objective, we also *directly* ask a (randomly) selected sample of our respondents about their experience with tax evasion, and compare the resulting incidence with that stemming from the list experiments. The difference between these two sets of responses measures the extent of social desirability bias, which reflects how willing respondents are to admit to engaging in the sensitive behavior when asked directly as opposed to indirectly. Surprisingly, our sample shows no significant social desirability bias. Since social desirability bias is plausibly nonrandom (Gonzalez-Ocantos, de Jonge, Meléndez, Osorio, & Nickerson, 2012), we also verify that the absence of bias holds overall as well as across a number of individual characteristics. We can thus confidently use answers to direct questions to examine evasion. Our third and final objective is to provide an overview of the resulting correlates for tax evasion, comparing our findings with those reported in the literature. To do this, we rely on a sensitivity analysis that helps us identify robust correlates that are not sensitive to model selection.

Our findings indicate that individuals with less favorable views about the rule of law (those who believe that bribing, resorting to violence, or taking justice into their own hands may be justified, as well as those who feel that it may be acceptable for authorities to violate the law in order to capture criminals) have a higher incidence of VAT evasion. Some of these correlations, in particular acquiescing to bribery, correspond to those found in the literature on “tax morale” that emphasizes the costs of tax evasion beyond strict economic cost–benefit calculations. Negatively reciprocal citizens (those who like to seek revenge on others for wrongdoings) also evade taxes more frequently, which may suggest that they respond to an inefficient state by not paying taxes; interestingly, we find no correlation with positively reciprocal individuals (those wanting to return favors to others).

Multiple dimensions of individual wealth and income – including household assets or land ownership, the per capita value of household expenditures, and not being affected by negative economic shocks – are correlated with lower levels of VAT evasion. Education also correlates negatively with evasion. We also find an independent negative correlation between VAT evasion and the extent of informality and the average size of businesses in households’ communities. We find that the physical presence of the state is associated with less tax evasion, and along these lines people who live in places with a history of higher levels of guerrilla activity (and thus plausibly lying beyond the state’s control) also evade more. Likewise, belief in democracy (defending elections as the preferred means of selecting

political leaders) is negatively correlated with tax evasion. Finally, of the other variables not robustly associated with evasion, one particularly interesting set concerns questions that capture individual beliefs about how actively the government should be involved in their lives (including, in particular, whether the state should have aggressive policies against inequality and whether individual welfare should depend on individuals themselves, or on the state).

Further examining the underlying forces of these empirical patterns is particularly relevant, as the theoretical and empirical literatures on this topic are relatively under-developed, despite the relevance of indirect taxation and evasion worldwide. Indeed, in many countries with prevailing poverty, high income inequality, and weak tax administration, indirect taxation (and VAT in particular) is an important source of revenue (De Jantscher, 1986; Besley & Persson, 2014). Moreover, while VAT was often introduced partly to cope with evasion and administration difficulties, evasion of this tax is widespread. Our estimates for Colombia indicate that nearly one out of every five people regularly avoid the VAT when purchasing goods. Although calculated using a completely different methodology, this share of VAT evaders is roughly in line with estimates of the amount of tax evaded in Colombia, at somewhat more than 20% of potential collection.<sup>3</sup> While the extent of VAT evasion varies around the world, figures for other countries are similarly significant. In Latin America, it is typically widespread, with estimates reaching 21.2% in Argentina in 2006, 11.0% in Chile in 2005, and 20% in Mexico in 2006 (Gómez-Sabaini & Jiménez, 2011).

Our dataset can help move the literature forward given the richness of the microeconomic information available. Indeed, the literature on the hidden economy (and particularly tax evasion) can follow a macro or a micro approach (Gemmell & Hasseldine, 2012).<sup>4</sup> The macro approach compares actual and potential revenue to determine the overall extent of evasion. These comparisons rely on extensive behavioral assumptions about what would have happened under stricter enforcement or additional revenue, when of course higher taxes and enforcement might change the structure of the economy and the behavior of different actors, including tax compliers and tax evaders. Moreover, in the macro approach it is inherently difficult to study these potential behavioral responses, since there is no indication of the type of people who are more likely to evade.

These limitations may be partly overcome by approaches such as ours, since taxpayer surveys help identify the determinants of evasion or undeclared income given the richness of demographic information (Andreoni, Erard, & Feinstein, 1998, p. 837). The main disadvantage, however, is social desirability bias (Slemrod, 2007). This discussion helps underscore

---

<sup>3</sup> Avila and Cruz (2007) show that VAT evasion declined from 31.8% in 1998 to 23.5% in 2006. Parra and Patiño (2010) report a figure of 20% for 2008, which is roughly in line with Corbacho, Fretes, and Lora (2013), who estimate it slightly below 25% in 2008 and slightly above 25% in 2010.

<sup>4</sup> For a discussion, see also Khlif and Achek (2015).

the unique advantages of our data. We have, in effect, the best of both worlds: (1) a very comprehensive survey with a wealth of information from survey respondents to examine the drivers of this sensitive behavior and (2) direct evidence that responses are not hampered by reporting biases. While list experiments have been used to examine a number of sensitive behaviors, to the best of our knowledge ours is the first paper examining tax evasion. Our analysis illustrates some of the possibilities of exploiting our data by analyzing the main correlates of tax evasion, verifying those that are not sensitive to model selection. Moreover, a key advantage of the method as implemented in this paper is the direct estimation of the extent of social desirability bias, a magnitude of interest in itself as it reveals the extent to which citizens internalize that tax evasion is an illegal and/or socially undesirable behavior. The absence of social desirability bias in our sample suggests that these concerns are not sufficiently embedded in Colombian society, and this may be one key obstacle for effective tax collection.

The paper proceeds with a brief examination of the empirical methods. Section 2.1 describes the key questions and validates the underlying assumptions, and Section 2.2 explains the “extreme bounds methodology,” a sensitivity analysis approach we implement to assess the robustness of the correlations in our data. Section 3 reports the main results on the incidence VAT evasion and documents the absence of social desirability bias in our survey responses. Section 4 presents the main findings, using the existing literature as a guide to study the most salient features of the data, and describing the robust correlates of evasion. Section 5 concludes.

## 2 Empirical methods

### 2.1 List experiments

To measure VAT evasion, respondents are randomly assigned to various groups. In the *Treatment* group, households are told “I will read a list of five (5) actions that people follow to save money when shopping. I want you to tell me how many of these five things you do regularly. Do not tell me WHICH, ONLY HOW MANY.” Then respondents are handed a card with the following options:

1. You choose the cheapest brand even if it is of lower quality,
2. You wait for sales on the best brands,
3. You buy in cheaper outlets even if they are far from your home,
4. **You accept buying without a receipt, to avoid paying VAT,**



5. You buy in bulk.

In the first control group (*Control 1*), respondents are confronted with a similar prompt and list, except the sensitive item (marked in bold above for emphasis, but not on the list used in the survey) is not in the list. Following the application of the list experiment, those in *Control 1* (who have not seen the sensitive item) are asked directly: “Could you tell me if you normally accept buying without a receipt, to avoid the VAT.” A third group, *Control 2* was not presented with the list experiment; respondents in this group were *only* asked this question directly, in case observing the control list biases responses.<sup>5</sup>

The premise of the experiment is that when asking indirectly by using a list, individuals are willing to answer truthfully even if social norms suggest that there is a “correct” answer. Since respondents in the *Treatment* condition only differ from those in *Control 1* in that they are presented with tax evasion as an option in the list, the difference in the number of actions reported by the two groups estimates the proportion of individuals that regularly evades the VAT.

The key assumption that individuals responding to the treatment and control questions are similar must hold in order for our exercises to be valid. While this should be the case by design since the groups were assigned randomly, we used the baseline survey from 2010 to verify balance on a number of observable baseline characteristics. Also, since randomization could fail in the field, we double checked balance on covariates in the follow-up 2013 survey when the list experiments were conducted. Online Appendix Tables A-2 to A-3 show that there are no systematic differences between the treatment and control groups, suggesting that their composition is unlikely to contaminate our results.

Additional assumptions underlie the ability of the list experiments to faithfully capture the behaviors of interest (Blair & Imai, 2012). The first main assumption is that there are no *design effects* – i.e. that the addition of a sensitive item does not affect the response on the sum of control items. The second assumption is that respondents answer truthfully to the sensitive item (i.e., that there are no liars). With these two assumptions, the difference in means between the treatment and control groups is an unbiased estimator of the incidence of the sensitive items.

Table 1 takes a first broad view of the results for the treatment and control lists. The average number of actions taken to save money in purchases is 1.68 in the control and 1.82 in the treatment group, which produces an estimated incidence of tax evasion of 13.68%.

This simple difference in means can also be computed for subsets of the population to study possible correlates of the sensitive behavior. However, this is not statistically

---

<sup>5</sup> Since we do not find significant differences in responses to the direct question between *Control 1* and *Control 2* in either experiment, we include both groups throughout when examining the direct questions.

**Table 1: Response frequency for treatment and control lists**  
**Tax evasion**

	(1)	(2)	(3)	(4)
	Control group		Treatment group	
Response value	Frequency	Proportion (%)	Frequency	Proportion (%)
0	120	4.5	123	4.1
1	1,129	42.1	1,165	38.5
2	1,009	37.6	1,098	36.3
3	330	12.3	454	15.0
4	95	3.5	144	4.8
5			42	1.4
Average	1.68		1.82	

**Notes:** The table lists the frequency in the number of actions people take to save money in purchases. The treatment list includes the same options as the corresponding control list, plus the following sensitive item: “You accept buying without a receipt, to avoid paying the VAT.”

efficient, and Imai (2011) and Blair and Imai (2012) propose maximum likelihood estimators to efficiently explore the role of correlates, not just of the sensitive behavior but also of the extent of social desirability bias in survey responses. We follow these approaches (explained in greater detail in the companion paper on clientelism Fergusson, Molina, and Riaño (2017)) in our empirical investigation. Finally, the ‘no design’ and ‘no liar’ assumptions can also be tested following Blair and Imai (2012). In Appendix A.3, we fail to reject the null hypothesis of no design effects, as well as of the two most common sources of untruthful answers: ceiling and floor effects.

We next explain the method we use to identify the most robust correlates of tax evasion in our data.

## 2.2 Uncovering robust correlates

To systematically establish which variables are most robustly correlated with evasion, we use the extreme bounds methodology (Leamer, 1985). The method (also explained in more detail in Fergusson, Molina, and Riaño (2017)) proceeds by estimating different models for the outcome of interest on a key covariate of interest and (various permutations) of additional controls. This then produces an entire distribution of estimated coefficients  $\beta_j$  for the key covariate of interest. Sala-i Martin (1997) proposes finding the cumulative density function of this distribution to the left and right from zero. The largest of these two (hereafter  $CDF(0)$ ) is the proportion of interest because it indicates where the coefficient is concentrated. Such densities can be recovered from the mean and standard deviation of the coefficient’s distri-

bution assuming normality, and weights proportional to some goodness-of-fit measure (like the adjusted R-squared or the integrated likelihood) can be used when computing such moments. However, with endogenous covariates the unweighted version may be preferable since endogenous regressions will have a better fit. The normality assumption can also be relaxed, computing  $CDF(0)$  for each regression, and only then finding the (weighted) average  $CDF(0)$ .

In this approach, variables that appear to be “significantly” correlated with the outcome are those with a (weighted)  $CDF(0)$  larger than 0.95, or another benchmark confidence level.<sup>6</sup> In our exercise, we focus on the average (weighted and unweighted) coefficient, and the cumulative density (both assuming and relaxing normality).<sup>7</sup>

### 3 Incidence and (no) social desirability bias

According to Colombia’s national tax agency (*Dirección de Impuesto y Aduanas Nacionales*, DIAN), the taxes collected in 2015 amounted to about 15% of the country’s GDP,<sup>8</sup> and despite an increase of at least 6 percentage points since 1990, this percentage is still slightly lower than the Latin American average (*Gómez-Sabaini & Jiménez, 2011*). We focus on VAT, which is the most important indirect tax (representing 28.69% of all government revenues in 2015)<sup>9</sup> and the one that most respondents are familiar with. Corporate taxes are of course limited to formal entrepreneurs, while direct taxes like income and property taxes are reserved for the relatively wealthy but also particularly weak in Colombia (*Comisión de expertos para la equidad y la competitividad tributaria, 2015*). Moreover, VAT evasion is an important concern in Colombia; DIAN campaigns have urged consumers to demand a receipt and pay the VAT. Thus our tax evasion question enquires about a practice that is familiar enough to respondents that they can realistically reflect on whether they are willing to engage in it to save on their purchases.

Figure 1 plots the basic incidence of tax evasion and reveals the absence of social desir-

---

<sup>6</sup>We warn, however, that though the method helps us uncover the correlations that are not sensitive to model selection, we remain cautious by not providing causal interpretations of our findings.

<sup>7</sup>We also report Leamer’s extreme bounds at the 95% level. The lower extreme bound is simply the lowest value of  $\beta_j - \tau\sigma_j$ , and the upper extreme bound is the largest value of  $\beta_j + \tau\sigma_j$ , with  $\tau$  the critical value for the confidence level and  $\sigma_j$  the standard error for  $\beta_j$ . If both bounds have the same sign then the corresponding variable is a robust correlate of tax evasion. However, this criteria is overly conservative, potentially declaring a correlation fragile on the basis of a single model.

<sup>8</sup> DIAN, Estadísticas Cifras de la gestión y logros de la entidad, Recaudo de los tributos administrados por la DIAN, Available at <http://www.dian.gov.co/contenidos/cifras/estadisticas.html>, last accessed on February 24, 2017.

<sup>9</sup> Compared to the most developed countries in Latin America, VAT in Colombia during 2010 represented 5.3% of GDP, lower than Argentina (8.1%), Brazil (13%), Chile (8.1%) and higher than Mexico (3.8%) (*Corbacho et al., 2013*).

ability bias.<sup>10</sup> When asked directly, 19.3% (SE = 0.005) of respondents reported that they had purchased an item without receiving a receipt to avoid paying VAT. The incidence of avoidance was higher in rural areas (21.7%) than in urban areas (17.2%). Numbers from the list experiment are very close, and the point estimate for social desirability bias is again very small and not statistically significant (-1.3 percentage points in urban areas and -1.2 percentage points in rural areas). Table 2 investigates the possible determinants of social desirability bias and finds that only a handful of individual traits are significant; in these cases, the bias is typically small and the point estimates are usually negative. We conclude that respondents are willing to openly report their tax evasion behavior. One possible explanation for this is that respondents in our survey are comfortable enough with survey officers, having been visited by the organization three years earlier for the baseline survey and answering a long questionnaire, to provide honest answers. It could also be the case that this reflects broad acceptance of evasion in Colombia more generally. In Fergusson, Molina, and Robinson (2017) we argue precisely that this reflects the nature of a political equilibrium in which clientelism and a weak state reinforce each other as normal features of the political landscape. This hypothesis is also consistent with citizens openly reporting whether they sell their vote (Fergusson, Molina, & Riaño, 2017).

## 4 Correlates

In this section we present the main correlates of tax evasion implementing the methodology described in Section 2.2.<sup>11</sup> To guide our inclusion of relevant variables, we briefly review and discuss the most relevant literature. Table 3 reports descriptive statistics. Definitions and sources for the latter are in Table A-1. To facilitate identifying the magnitude of the correlations, we standardize all variables in the regression analysis.

The empirical evidence on tax evasion is surprisingly thin. The seminal economic model of tax evasion (Allingham & Sandmo, 1972) and its successors produce a number of theoretical predictions, but these are hard to test given the inherent difficulties in measuring tax evasion. As Slemrod and Yitzhaki (2002, p. 1440) explain, citing Harvey Galper, “Regression analysis of tax evasion is straightforward, except for two problems: you can’t measure the left-hand side variable, and you can’t measure the right-hand side variables!”

In addition, the predictions on some key variables (such as the tax rate or the con-

---

<sup>10</sup> Estimates control for the set of individual characteristics in Table 2, though average incidence is not sensitive to changes in these controls.

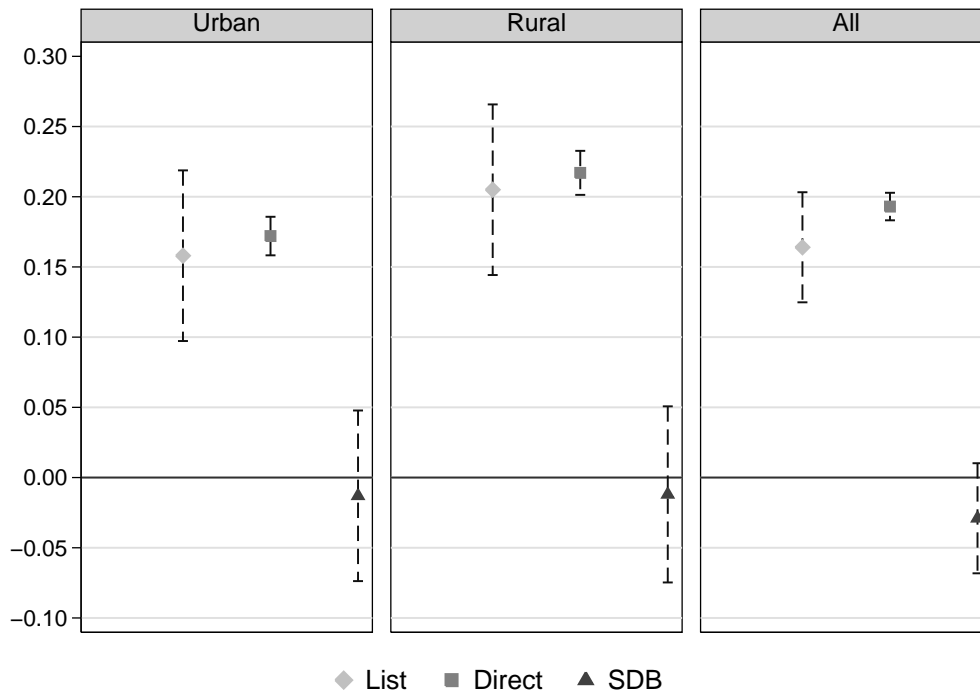
<sup>11</sup> Appendix A.4 also reports simple alternative bivariate and multivariate regressions, which produce similar conclusions.

**Table 2: No social desirability bias: tax evasion**

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
<i>Variables</i>		Urban sample			Rural sample			Full sample		
		<i>List</i>	<i>Direct</i>	<i>SDB</i>	<i>List</i>	<i>Direct</i>	<i>SDB</i>	<i>List</i>	<i>Direct</i>	<i>SDB</i>
Age	(18,40]	0.149*** (0.041)	0.181*** (0.012)	-0.032 (0.043)	0.254*** (0.058)	0.226*** (0.014)	0.028 (0.060)	0.184*** (0.033)	0.200*** (0.009)	-0.017 (0.035)
	(40,55]	0.135*** (0.046)	0.179*** (0.011)	-0.043 (0.047)	0.154*** (0.042)	0.207*** (0.012)	-0.053 (0.043)	0.128*** (0.027)	0.192*** (0.008)	-0.063** (0.029)
	>55	0.180*** (0.065)	0.143*** (0.014)	0.037 (0.066)	0.198*** (0.056)	0.222*** (0.015)	-0.023 (0.058)	0.182*** (0.041)	0.186*** (0.011)	-0.004 (0.043)
Education	Primary or less	0.200*** (0.053)	0.185*** (0.013)	0.014 (0.055)	0.214*** (0.034)	0.228*** (0.009)	-0.014 (0.036)	0.202*** (0.028)	0.214*** (0.008)	-0.012 (0.029)
	Secondary	0.110*** (0.040)	0.181*** (0.011)	-0.071* (0.041)	0.136** (0.059)	0.185*** (0.016)	-0.049 (0.061)	0.102*** (0.030)	0.181*** (0.009)	-0.079** (0.032)
	College	0.157*** (0.059)	0.138*** (0.013)	0.018 (0.061)	0.236 (0.277)	0.172*** (0.045)	0.064 (0.281)	0.141*** (0.054)	0.141*** (0.013)	0.001 (0.056)
Gender	Men	0.114*** (0.043)	0.164*** (0.012)	-0.050 (0.044)	0.185*** (0.042)	0.206*** (0.012)	-0.021 (0.044)	0.142*** (0.029)	0.186*** (0.008)	-0.044 (0.030)
	Women	0.171*** (0.036)	0.176*** (0.009)	-0.006 (0.037)	0.210*** (0.041)	0.227*** (0.011)	-0.017 (0.043)	0.173*** (0.026)	0.198*** (0.007)	-0.025 (0.027)
Employment	No	0.165*** (0.052)	0.173*** (0.013)	-0.008 (0.053)	0.189*** (0.051)	0.197*** (0.013)	-0.008 (0.052)	0.160*** (0.034)	0.185*** (0.009)	-0.025 (0.035)
	Yes	0.144*** (0.034)	0.171*** (0.008)	-0.027 (0.035)	0.203*** (0.039)	0.227*** (0.010)	-0.024 (0.040)	0.162*** (0.023)	0.197*** (0.006)	-0.035 (0.024)
Frequency of voting	Low	0.151*** (0.044)	0.189*** (0.013)	-0.038 (0.045)	0.213*** (0.067)	0.185*** (0.018)	0.027 (0.069)	0.151*** (0.033)	0.187*** (0.010)	-0.036 (0.035)
	High	0.151*** (0.038)	0.163*** (0.008)	-0.012 (0.039)	0.196*** (0.033)	0.224*** (0.009)	-0.028 (0.034)	0.164*** (0.023)	0.195*** (0.006)	-0.031 (0.024)
Vote for the same party	No	0.137*** (0.033)	0.169*** (0.008)	-0.032 (0.034)	0.192*** (0.046)	0.195*** (0.010)	-0.003 (0.047)	0.146*** (0.025)	0.179*** (0.007)	-0.033 (0.026)
	Yes	0.187*** (0.053)	0.180*** (0.014)	0.007 (0.055)	0.206*** (0.040)	0.242*** (0.012)	-0.036 (0.042)	0.186*** (0.031)	0.219*** (0.009)	-0.032 (0.033)
Secret ballot	No	0.169*** (0.046)	0.175*** (0.012)	-0.006 (0.047)	0.274*** (0.062)	0.251*** (0.016)	0.023 (0.064)	0.200*** (0.036)	0.207*** (0.010)	-0.007 (0.037)
	Yes	0.141*** (0.036)	0.170*** (0.009)	-0.030 (0.037)	0.170*** (0.033)	0.204*** (0.009)	-0.035 (0.035)	0.143*** (0.022)	0.187*** (0.006)	-0.045* (0.023)
Religion	None or other	0.168*** (0.056)	0.168*** (0.015)	0.000 (0.058)	0.363*** (0.083)	0.228*** (0.021)	0.135 (0.086)	0.236*** (0.047)	0.190*** (0.012)	0.046 (0.049)
	Catholic	0.147*** (0.034)	0.173*** (0.008)	-0.026 (0.035)	0.170*** (0.033)	0.215*** (0.009)	-0.046 (0.034)	0.143*** (0.022)	0.194*** (0.006)	-0.051** (0.022)
Skin color	White	0.135*** (0.033)	0.163*** (0.009)	-0.028 (0.035)	0.159*** (0.038)	0.211*** (0.010)	-0.052 (0.040)	0.137*** (0.024)	0.185*** (0.007)	-0.048* (0.025)
	Black	0.182*** (0.050)	0.190*** (0.013)	-0.008 (0.052)	0.266*** (0.050)	0.227*** (0.013)	0.039 (0.051)	0.204*** (0.033)	0.209*** (0.009)	-0.004 (0.035)
Shock	No	0.123** (0.055)	0.141*** (0.011)	-0.018 (0.056)	0.221*** (0.060)	0.190*** (0.015)	0.030 (0.062)	0.149*** (0.036)	0.162*** (0.009)	-0.013 (0.037)
	Yes	0.165*** (0.034)	0.188*** (0.009)	-0.023 (0.035)	0.190*** (0.035)	0.227*** (0.010)	-0.038 (0.036)	0.166*** (0.023)	0.207*** (0.007)	-0.041* (0.024)
Wealth	Below median	0.149*** (0.036)	0.217*** (0.011)	-0.068* (0.038)	0.202*** (0.038)	0.237*** (0.011)	-0.035 (0.039)	0.188*** (0.027)	0.227*** (0.008)	-0.039 (0.028)
	Above median	0.152*** (0.047)	0.131*** (0.009)	0.021 (0.048)	0.194*** (0.049)	0.194*** (0.011)	0.001 (0.050)	0.129*** (0.028)	0.154*** (0.007)	-0.025 (0.029)

**Notes:** For each area, rural, urban, and the full sample, the table shows the average incidence of tax evasion using the *list* experiment (with methods as described in Section 2.1), the *direct* question (using a logit model), and the difference between the two or social desirability bias  $-SDB$ . Standard errors are computed using Monte Carlo simulations, and estimations control for the set of variables listed in the table. \* is significant at the 10% level, \*\* is significant at the 5% level, \*\*\* is significant at the 1% level. See Table A-1 for a description of all variables.

**Figure 1: Incidence and social desirability bias: Tax evasion**



**Notes:** For each area, the figure shows the incidence of VAT evasion as implied by the list experiment (diamond), direct question (square), and the difference between these two measures, capturing the extent of social desirability bias (SDB, triangle). Lines mark 95% confidence bounds. Estimates in this figure control for the set of individual characteristics listed in Table 2.

tributors’ income levels) are often ambiguous, depending on the modeling assumptions.<sup>12</sup> Empirical analyses should thus ideally make a careful connection to theoretical predictions, yet [Andreoni et al. \(1998, p. 836\)](#) complain that this has not been the case: “recent empirical work is... only loosely connected with theory... and, partly as a result, few of the recent empirical findings have led to productive new theorizing.”

Moreover, most theoretical and empirical analyses (in both observational and experimental settings) focus on income tax evasion. Several authors have pointed out that indirect tax evasion has been neglected by comparison ([Marrelli, 1984](#); [Virmani, 1989](#); [Matthews & Lloyd Williams, 2001](#)).<sup>13</sup> As discussed before, and as our data reveals, this is unfortunate

<sup>12</sup> For instance, one key aspect is whether the tax authority is modeled as exogenously determining enforcement or as playing a game with taxpayers. This is also important empirically: it reminds us that models in which tax policy variables are taken as exogenous can be misspecified ([Andreoni et al., 1998](#)).

<sup>13</sup> The main surveys of the literature clearly reveal this bias. In the recent reviews of [Khlif and Achek \(2015\)](#) and [Hanlon and Heitzman \(2010\)](#), VAT is not discussed. In the survey on tax compliance by [Andreoni et al. \(1998\)](#) “VAT” is not mentioned once. In [Slemrod and Yitzhaki \(2002\)](#), VAT is only highlighted to note that it has been favored given the difficulty in avoiding it, incentives for self-policing, or as a complement to

**Table 3: Summary statistics**

	(1)	(2)	(3)	(4)	(5)	(6)
	Observations	Mean	Median	Std. Dev.	Min	Max
<b>Panel A. Dependent variable</b>						
Tax evasion	5,275	0.193	0.000	0.395	0.000	1.000
<b>Panel B. Covariates</b>						
Age	5,275	46.518	46.000	12.657	15.000	89.000
Agree with bribery	5,275	0.137	0.000	0.344	0.000	1.000
Authorities violate law	5,275	0.344	0.000	0.475	0.000	1.000
Catholic	5,275	0.814	1.000	0.389	0.000	1.000
Commerce sector	5,275	0.434	0.000	0.496	0.000	1.000
Education	5,275	0.458	0.000	0.498	0.000	1.000
Employment firms 2+	5,275	0.585	0.585	0.195	0.206	0.961
Evangelical/Pentecostal	5,275	0.127	0.000	0.333	0.000	1.000
Fractionalization	5,275	0.623	0.641	0.102	0.422	0.813
Gov. against inequality	5,275	0.935	1.000	0.247	0.000	1.000
Government role	5,275	0.887	1.000	0.316	0.000	1.000
Guerrillas	5,275	0.222	0.000	1.650	0.000	21.085
HH expenses	5,275	0.093	0.000	0.155	0.000	1.000
Homicide rate	5,275	27.039	18.349	25.681	0.000	163.159
Independent	5,275	0.159	0.000	0.366	0.000	1.000
Justice into own hands	5,275	0.295	0.000	0.456	0.000	1.000
Lands	5,275	0.034	0.000	0.180	0.000	1.000
Negative reciprocity	5,275	0.193	0.000	0.395	0.000	1.000
Neighbor cell phones	5,275	0.117	0.000	0.321	0.000	1.000
Neighbor loans	5,275	0.192	0.000	0.394	0.000	1.000
Other religion	5,275	0.023	0.000	0.149	0.000	1.000
Own welfare	5,275	0.952	1.000	0.215	0.000	1.000
Paramilitaries	5,275	1.469	0.000	7.309	0.000	68.367
Polarization	5,275	0.831	0.852	0.090	0.582	1.000
Pop. density	5,275	0.001	0.000	0.002	0.000	0.013
Popular vote	5,275	0.953	1.000	0.211	0.000	1.000
Positive reciprocity	5,275	0.972	1.000	0.165	0.000	1.000
Rural population	5,275	0.347	0.394	0.276	0.002	0.954
Shock	5,275	0.693	1.000	0.461	0.000	1.000
State presence	5,275	-0.010	0.590	1.881	-7.305	3.126
Use of violence	5,275	0.203	0.000	0.402	0.000	1.000
Wealth	5,275	-0.230	-0.583	2.635	-5.296	5.887
Win margin	5,275	0.159	0.131	0.113	0.001	0.476
Woman	5,275	0.587	1.000	0.492	0.000	1.000
Workers by firm	5,275	3.084	2.304	1.864	1.263	21.970

**Notes:** *Tax evasion* is a dummy variable that equals 1 if the respondent normally accepts buying items without a receipt, to avoid paying VAT. Covariates in Panel B, and their sources, are described in Table A-1. The sample of respondents is the set of individuals in groups *Control 1* and *Control 2* as described in Section 2.1, namely, those asked directly about tax evasion.

since indirect taxation (and VAT evasion in particular) is a pressing issue in many countries.

Undoubtedly, if the empirical literature is lagging behind on analyses of direct tax evasion, which has attracted the most attention, the situation is worse for indirect taxation. We thus briefly discuss the potential determinants that we can most easily approximate with our data. To guide our analysis of some suggestive correlations, we rely on the available literature (even if it refers to income tax) and attempt to draw useful lessons for indirect taxation. However, more needs to be done to understand the determinants of indirect tax cross information to raise income taxes (Alt, 1983).

evasion, both theoretically and empirically. Our micro data provide a unique opportunity to move forward, combining rich individual data with direct evidence of no reporting biases.

## 4.1 Indirect tax evasion

We start with a brief discussion of key insights from the comparatively limited literature on indirect tax evasion in order to illustrate the potential of our data.<sup>14</sup> Below, we discuss the determinants of tax evasion that may apply to any type of tax obligations (or that were designed to analyze income tax evasion but may also apply to indirect taxation).

Indirect tax evasion has focused on firm behavior, at least since the early work of [Marrelli \(1984\)](#), which many consider to be the seminal contribution ([Virmani, 1989](#); [Yaniv, 1988](#); [Arias, 2005](#)). This emphasis likely reflects the fact that firms (not consumers) are typically accountable to the tax authorities for reporting and paying indirect taxes. However, since there is typically some degree of collusion between the buyer and seller, features of both may be important in determining the extent of indirect tax (particularly VAT) evasion.<sup>15</sup> Some argue that buyers always prefer the cheaper “tax free” option, so the only relevant side is the firm.<sup>16</sup> Yet the consumer decision to avoid paying VAT is not necessarily costless: it may involve relinquishing some benefits (such as a warranty on the product, which is only valid with a receipt), facing risks (like carrying more cash, to more easily conceal the evasion), or paying a psychological or social moral cost (as the literature on tax morale discussed below suggests).

Models of indirect tax evasion produce quite divergent results depending on the specific assumptions ([Arias, 2005](#)). In [Marrelli \(1984\)](#) and others, the key question is the extent to which a monopolistic firm under-reports income on sales in order to evade taxes. In addition to the expected result that harsher penalties – and an increase in the probability of getting caught – increase the declared tax base, two key predictions are that larger firms declare a greater fraction of the tax base, and that changes in the tax rate have ambiguous

---

<sup>14</sup> Key references in the literature include [Marrelli \(1984\)](#); [Wang and Conant \(1988\)](#); [Yaniv \(1988\)](#); [Virmani \(1989\)](#); [Cremer and Gahvari \(1993\)](#); [Yaniv \(1995\)](#).

<sup>15</sup> [Yaniv \(1988\)](#) studies tax withholding more generally, particularly firms’ decisions to under-report the tax withheld from employees, and finds that withholding makes it hard for employees and firms to evade taxation without colluding. But a firm may still risk remitting to the government less than the amounts withheld, especially for wage earners who are not required to file an income tax return (which the tax authorities could use for comparison). VAT evasion creates similar interactions between the firm and its buyers.

<sup>16</sup> VAT has been said to facilitate enforcement through an incentive structure that creates a third-party reported paper trail on transactions between firms. But this is not the case where sales are made to final consumers, who (unlike firms in the chain) do not need a receipt to deduct input costs from their VAT bill. A field experiment by [Pomeranz \(2015\)](#) of Chilean firms and consumers is consistent VAT evasion in final sales.



effects. Finally, when comparing the indirect tax to a profit tax of equal yield, the former is evaded to a lesser extent with decreasing risk aversion. The literature has examined a number of variations of this benchmark model, among which [Arias \(2005\)](#) highlights the following: market competition (competitive, monopolistic), tax function (ad valorem, specific, withholding, profit), cost functions, attitudes toward risk (averse, neutral), probability of detection function (fixed, variable), and tax evasion as a percentage or an amount.

[Gordon \(1990\)](#) complements this analysis by noting how “under the counter” (discounted) cash sales produce unrecorded income, thus facilitating income tax evasion, which may also interest firms. Hence, cash sales provide a link between direct and indirect taxation, which Gordon describes as particularly prevalent in the services sector. Some empirical papers use macro estimates of tax evasion at the sector level and find, consistent with [Gordon \(1990\)](#), more evasion in the restaurant/take-out and hairdressing sectors than in the clothing/footwear and furniture/floor-covering sectors ([Matthews & Lloyd Williams, 2001](#)). [Zídková \(2014\)](#), using data from 24 EU member states, however, finds a smaller VAT gap in countries with larger restaurant and hotel services sectors, which she associates more generally with tourism.

This literature invites an examination of the characteristics of the consumer economic environment (e.g., are consumers likely to interact with larger or smaller firms? Or firms in certain sectors?). [Table 4 \(row 9\)](#) illustrates that respondents living in municipalities where more formal employment is in “larger” establishments (i.e., those with at least two employees) are significantly and negatively correlated with tax evasion. Similarly, tax evasion is lower where more workers are formally employed ([row 18](#)). The correlation with the importance of the commerce sector ([row 22](#)), while positive, is somewhat less robust.<sup>17</sup>

Below we examine the correlation between buyers who avoid paying VAT to save on expenses and household and individual features.

## 4.2 Institutional and contextual factors

In the classical model of tax evasion ([Allingham & Sandmo, 1972](#)), citizens who under-report their income do so at the risk that the government will find out and impose a penalty, which leads to the prediction that equilibrium evasion decreases with the likelihood of an audit and the size of the penalty. Research on income tax evasion has found support for both predictions ([Andreoni et al., 1998](#); [Christie & Holzner, 2006](#); [Fortin, Lacroix, & Villeval, 2007](#); [Coricelli, Joffily, Montmarquette, & Villeval, 2010](#)). Where information about the likelihood of an audit and/or punishment is not available, measures of state capacity are frequently

---

<sup>17</sup>To easily identify the most important correlates, variables are sorted from most to least robustly (or “significantly”) correlated with tax evasion.

**Table 4: Tax evasion  
Robust correlates**

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Leamer's		Coefficient		Standard error		CDF(0)		
	Lowest	Upperest	Weighted	Unweighted	Weighted	Unweighted	Weighted*	Weighted	No weighted
1. Wealth	-0.175	-0.037	-0.114	-0.114	0.023	0.023	1.000	1.000	1.000
2. Use of violence	0.018	0.105	0.071	0.071	0.014	0.014	1.000	1.000	1.000
3. Lands	-0.067	-0.015	-0.044	-0.044	0.010	0.010	1.000	1.000	1.000
4. Agree with bribery	0.019	0.096	0.064	0.064	0.014	0.014	1.000	1.000	1.000
5. Guerrillas	0.021	0.108	0.067	0.067	0.014	0.014	1.000	1.000	1.000
6. Justice into own hands	0.011	0.099	0.067	0.067	0.014	0.014	1.000	1.000	1.000
7. Negative reciprocity	0.009	0.101	0.064	0.064	0.015	0.015	1.000	1.000	1.000
8. Rural population	-0.033	0.173	0.081	0.081	0.024	0.024	0.997	1.000	1.000
9. Employment firms 2+	-0.167	0.031	-0.080	-0.080	0.024	0.024	0.996	0.999	0.999
10. Education	-0.087	0.009	-0.041	-0.042	0.013	0.013	0.996	0.999	0.999
11. Other religion	-0.079	0.001	-0.027	-0.027	0.011	0.011	0.994	0.994	0.994
12. Win margin	-0.042	0.131	0.047	0.047	0.018	0.018	0.993	0.995	0.995
13. State presence	-0.076	0.009	-0.036	-0.037	0.015	0.015	0.991	0.993	0.993
14. Shock	-0.003	0.054	0.027	0.027	0.012	0.012	0.987	0.988	0.988
15. HH expenses	-0.072	0.017	-0.036	-0.036	0.015	0.015	0.984	0.994	0.994
16. Popular vote	-0.058	0.003	-0.028	-0.028	0.014	0.014	0.979	0.980	0.980
17. Authorities violate law	-0.022	0.064	0.031	0.032	0.014	0.014	0.976	0.989	0.990
18. Workers by firm	-0.124	0.059	-0.054	-0.054	0.019	0.019	0.967	0.997	0.997
19. Evangelical/Pentecostal	-0.121	0.031	-0.022	-0.022	0.013	0.013	0.957	0.962	0.962
20. Fractionalization	-0.152	0.131	-0.034	-0.034	0.018	0.018	0.944	0.966	0.966
21. Catholic	-0.109	0.057	0.017	0.017	0.013	0.013	0.896	0.901	0.899
22. Commerce sector	-0.037	0.072	0.019	0.019	0.016	0.016	0.859	0.879	0.879
23. Own welfare	-0.043	0.017	-0.014	-0.014	0.013	0.013	0.857	0.858	0.858
24. Age	-0.060	0.020	-0.014	-0.014	0.013	0.013	0.856	0.870	0.870
25. Independent	-0.018	0.042	0.013	0.013	0.013	0.013	0.844	0.846	0.847
26. Woman	-0.020	0.056	0.016	0.016	0.016	0.016	0.842	0.844	0.843
27. Pop. density	-0.052	0.099	0.003	0.002	0.017	0.017	0.752	0.562	0.554
28. Polarization	-0.153	0.113	0.005	0.005	0.019	0.019	0.751	0.603	0.606
29. Neighbor cell phones	-0.026	0.048	0.010	0.009	0.015	0.015	0.728	0.730	0.730
30. Neighbor loans	-0.044	0.032	-0.006	-0.007	0.014	0.014	0.674	0.674	0.675
31. Homicide rate	-0.079	0.056	0.004	0.005	0.020	0.020	0.672	0.590	0.592
32. Gov. against inequality	-0.039	0.029	-0.006	-0.006	0.015	0.015	0.646	0.646	0.646
33. Paramilitaries	-0.043	0.035	-0.004	-0.004	0.014	0.014	0.618	0.612	0.611
34. Positive reciprocity	-0.029	0.032	0.003	0.003	0.014	0.014	0.588	0.587	0.588
35. Government role	-0.036	0.036	0.002	0.002	0.014	0.014	0.565	0.542	0.544

**Notes:** The table reports statistics based on the extreme bounds methodology described in Section 2.2. The dependent variable of interest is a dummy indicating whether the respondent normally accepts buying items without a receipt, to avoid paying VAT. The independent variable of interest is indicated in each row. In each case, a sensitivity analysis is conducted by also including all possible permutations of up to three of the additional covariates listed in the table. Region fixed effects are always included, and standard errors are clustered at the community level. Lowest and Upperest are Leamer's lowest and upper extreme bounds for the coefficient of interest (that is, for the resulting coefficient of the variable listed in each row) at the 95% confidence level. The adjusted R-squared is used to weight statistics where indicated. The  $CDF(0)$  measures the largest accumulated density of the resulting coefficients of interest to the left or right of zero (whichever is largest). Column 7 reports the weighted  $CDF(0)$  assuming no normality, while Column 8 (9) reports the weighted (unweighted)  $CDF(0)$  assuming normality. See Table A-1 for a description of all variables.

used as proxies.<sup>18</sup> But state capacity may indicate more than just the audit probability; it also contributes to the general control of taxpayers: more efficient tax administration and information facilitates compliance (Kirchler, 2007). Political uncertainty and instability may

<sup>18</sup> For example, Christie and Holzner (2006) apply a *judicial/legal effectiveness index* to a panel of 29 European countries between 2000 and 2003, and Picur and Riahi-Belkaoui (2006) find that better legal systems are associated with lower levels of tax evasion.

also affect levels of tax avoidance: individuals may be more willing to hide their funds (or transactions) from the tax authorities when there is more uncertainty about the tax policies of future governments (Katz & Owen, 2013).

When we explore measures of state capacity and political uncertainty in our data we find some support for these basic predictions. A measure of municipality state presence (Table 4, row 13) is indeed negatively and very robustly correlated with VAT evasion. Examining violence and the crime rate is another way to gauge the role of state effectiveness and capacity, and we find that areas with more guerrilla activity (row 5) have higher levels of tax evasion. Yet neither the presence of paramilitaries (row 33) nor the homicide rate (row 31) correlates robustly with evasion.<sup>19</sup> Political competition, however, is negatively and robustly correlated with evasion (rows 12 and 20). The fractionalization index based on electoral returns correlates negatively with evasion and positively with the win margin. This finding could be interpreted as reflecting the role of political uncertainty and thus conflicting with the prediction in Katz and Owen (2013), yet it may also capture the influence of other factors measured by this variable, such as a healthier local democracy.<sup>20</sup> Indeed, tax evasion is less prevalent among respondents who believe that it is important that leaders are elected by popular vote (row 16).

The existence (and size) of an underground economy is also important, as it makes it easier to avoid paying taxes (Slemrod & Yitzhaki, 2002). A perhaps not surprising yet related finding is that income tax evasion is common among the self-employed (Fiorio & D’Amuri, 2006; Kirchler, 2007). Self-employment may facilitate cheating for several reasons, including less access to information sources relative to employees for tax authorities, a higher proportion of cash transactions, and, especially in developing countries, a higher incidence of informality that reduces both records for the government and taxpayers’ perceived benefits from the government, thus discouraging compliance (Slemrod, 2007). While this focus is mostly relevant for income tax evasion, informality can also influence VAT evasion via individual perceptions about perceived benefits and monitoring (transactions are not registered in informal markets, and, as our tax evasion question shows, not producing a receipt is often an essential component of noncompliance). Zídková (2014) finds a positive relationship between the VAT gap and the size of the shadow economy as a percentage of GDP. This is in line with the negative correlation in our data with formal employment in the municipality (row 18). For individual self-employment, we also find a positive (but weak) correlation on

---

<sup>19</sup> Informality and enforcement in general may also be weaker in rural areas, and indeed there is a positive and robust correlation, in row 8, between these variables and tax evasion. More densely populated areas exhibit less evasion, on average, but this correlation is not robust (row 27).

<sup>20</sup> The polarization index (row 28), which may be more relevant for uncertainty, is instead positively correlated with evasion, yet the coefficient is not significant.

average (row 25); the  $CDF(0)$  is under 95%.

### 4.3 Tax morale

The economic psychology of tax evasion emphasizes that cheating may have costs beyond the monetary consequences, that subjective perceptions may play a larger role than objective probabilities, that individual and social norms are relevant factors, and that individual behavioral characteristics such as motivation and behavior control are important (Kirchler, 2007). This has inspired a very broad literature on “tax morale” (Slemrod, 1998), which Luttmer and Singhal (2014) broadly define as all nonpecuniary motivations for tax compliance and factors outside the standard, expected utility framework, including: intrinsic motivation to pay taxes, feelings of guilt or shame, reciprocal motivations, influence of peer behavior, cultural or social norms interacting with these motivations, and deviations from standard expected utility maximization, such as loss aversion.

Tax morale thus includes a number of determinants operating through different mechanisms, many of which have been tested empirically, often experimentally.<sup>21</sup> Previous studies have shown that guilt and shame are particularly important and may improve the fit of the baseline – purely economic – model. Taxpayers may anticipate guilt from under-reporting despite escaping detection and experience shame if caught, which in turn influences their behavior (Andreoni et al., 1998; Erard & Feinstein, 1994; Grasmick & Bursik Jr, 1990).<sup>22</sup> While we do not have information on variables such as guilt or shame, we do have a number of individual responses on moral views and beliefs that relate to some of these hypotheses. Citizens with certain views evade more: those who agree that sometimes it is necessary to pay a bribe (row 4), think that some times authorities should violate the law to capture criminals (row 17), agree that using violence is sometimes justified (row 2), and agree with taking justice into their own hands (row 6). Thus, individuals who are more open to accepting anti-social or illegal behavior are more prone to avoid paying taxes.

Since taxation is at the heart of the social contract between citizens and the government, beliefs about potential breaches of the deal may justify and encourage evasion (Kirchler,

---

<sup>21</sup> Some experiments that provide information to individuals attempt to exploit these mechanisms to increase collection. For instance, Blumenthal, Christian, Slemrod, and Smith (2001) send normative appeals to taxpayers, yet fail to affect aggregate tax compliance behavior. Instead, Bott, Cappelen, Sørensen, and Tungodden (2014) experiment by sending a moral appeal or information about the probability of evasion detection, and find a large positive impact in both cases (reported income almost doubled): the moral appeal affected the *amount* reported, while the probability of evasion detection increased the *incidence* of tax payment.

<sup>22</sup> Experimental evidence from Coricelli et al. (2010) reveals that the risk of exposure (which involves publicly displaying the evader’s picture) deters evasion and arouses subjects’ emotions. Evasion and monetary sanctions also stir respondents’ emotions before they submit their decision, yet the low fines encourage (rather than deter) evasion.

2007; Slemrod, 2007). Litina and Palivos (2016) model a potential vicious circle of political corruption and tax evasion: “corruption may corrupt” when politicians speculate and citizens evade taxes. This may be yet another reason why a stronger state or legal system, and better functioning institutions more generally, may facilitate compliance. These mechanisms seem consistent with the correlation reported above between evasion and municipal state capacity and guerrilla presence.<sup>23</sup> Also along these lines, if taxpayers believe the tax system (the code and/or its enforcement) is unfair, compliance may fall. As Andreoni et al. (1998) put it, “in psychological terms, an unfair tax system could lead people to ‘rationalize’ cheating” (p. 851), a prediction that is consistent with the findings in Richardson (2006). More generally, lower levels of trust in the government (Torgler, 2003; Richardson, 2008), the prevalence of corruption and a more inefficient bureaucracy (Picur & Riahi-Belkaoui, 2006), and the prevalence of crime and less individual economic freedom (Riahi-Belkaoui, 2004) may encourage cheating.

Taxpayers who feel cheated because they believe their tax dollars are not well spent may reciprocate by not fully complying with their obligations (Spicer & Lundstedt, 1976; Smith, 1992). Indeed, experiments suggest that subjects more willingly pay taxes when they receive benefits from a public good (Alm, Jackson, & McKee, 1992) and when they report higher levels of satisfaction with laws and the government (Webley, 1991). Complementary to this logic is the idea that individuals who are more reciprocal will be more responsive and engage less in evasion when they think the government will act in their interests and other citizens are paying taxes (Slemrod, 2007).

In our data, we find that respondents who exhibit negative reciprocity (row 7) are more likely to evade. This correlation is extremely robust, but is weak with positive reciprocity (row 34). Since there is generally a poor perception of the state in Colombia, this may reflect that more (negatively) reciprocal voters rationalize cheating. To look at this more directly, we interact the measures of reciprocity with the measure of municipal state capacity, expecting reciprocal citizens to reduce the level of evasion when the state is more present. Table 5 shows that this is the case with positive reciprocity, but not with negative reciprocity. Also related, tax evasion is less prevalent among respondents who believe that it is important for leaders to be elected by popular vote (row 16).<sup>24</sup>

Peer influence is another important factor. Citizens may find it less costly to under-report

---

<sup>23</sup> Neither the variables that capture beliefs on the role that the government (as opposed to the people) plays in determining individual welfare (“Government role” and “Own welfare”) nor the views on how actively the state must combat inequality (“Gov. against inequality”) is robustly correlated with evasion. This is perhaps surprising, as one view is that those who support a stronger role for the government might be more willing to pay taxes.

<sup>24</sup> We also interact this variable with the reciprocity measures to examine if they are especially sensitive, but they do not appear to be.

income if their peers are used to doing so (Gordon, 1989); some theories include a utility payoff from behaving honestly and adhering to the standard pattern of peers’ behavior (Myles & Naylor, 1996). Examining peer effects requires more careful study<sup>25</sup>, thus it is perhaps not surprising that simply including the number of neighbors’ cell phone numbers they know and the number of neighbors to whom respondents can reach out for loans (rows 29 and 30) are not robustly correlated with evasion.

**Table 5: Tax evasion  
Interaction terms**

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Leamer’s		Coefficient		Standard error		CDF(0)		
	Lowest	Upperest	Weighted	Unweighted	Weighted	Unweighted	Weighted*	Weighted	Unweighted
1. State presence*Positive rec.	-0.043	0.002	-0.022	-0.022	0.010	0.010	0.986	0.986	0.986
2. State presence*Negative rec.	-0.022	0.036	0.007	0.007	0.013	0.013	0.715	0.716	0.716
3. Popular vote*Positive rec.	-0.021	0.034	0.007	0.007	0.013	0.013	0.697	0.697	0.697
4. Popular vote*Negative rec.	-0.025	0.033	0.004	0.004	0.014	0.014	0.626	0.626	0.626

**Notes:** The table reports statistics based on the extreme bounds methodology described in Section 2.2. For each row, we run a set of regressions varying all possible groups of up to three covariates (see Table 4) in addition to the interaction listed in the row, the lower-order direct terms, and region fixed effects. The dependent variable of interest is a dummy indicating whether the respondent normally accepts buying items without a receipt, to avoid paying VAT. Standard errors are clustered at the community level. Lowest and Upperest are Leamer’s lowest and upper extreme bounds for the coefficient of interest (that is, for the resulting coefficient of the variable listed in each row) at the 95% confidence level. The adjusted R-squared is used to weight statistics where indicated. The *CDF*(0) measures the largest accumulated density of the resulting coefficients of interest to the left or right of zero (whichever is largest). Column 7 reports the weighted *CDF*(0) assuming no normality, while Column 8 (9) reports the weighted (unweighted) *CDF*(0) assuming normality. See Table A-1 for a description of all variables.

More broadly, social norms may influence behavior. McGee (2011) focuses on culture and religion, documenting different religions’ perspectives on the matter. Richardson (2006) finds no evidence that countries with higher percentages of Protestants, Catholics or Muslims have a different pattern regarding tax evasion, though evasion is less common in countries where people are more likely to self-identify as religious (Richardson, 2008). In our data, the results for Catholic (row 21), Evangelical/Pentecostal (row 19), and other religions (i.e., Mormons, Jews, Jehovah’s Witness, row 11) indicate that except for Catholics, believers are less prone to evade than atheists or agnostics, the excluded category. Since Catholics constitute the largest religious group (81.5% in our data), it is likely that there is considerable variation within this group.

<sup>25</sup> Fortin et al. (2007) propose a model with different social interaction effects, including both endogenous interactions (social conformity effects, i.e., deriving a psychic payoff from adhering to a pattern in the reference group) and exogenous interactions (like the fairness effects discussed above). They discuss the enormous empirical difficulties in disentangling these effects, given that there may also be social learning (e.g., learn less costly ways to evade taxes from peers) and correlated effects (since individuals in a given group have similar characteristics or share a similar environment). Moreover, Manski’s reflection problem (Manski, 1993) (i.e., the simultaneity in the behavior of interacting agents introduces a collinearity between the mean outcome of the group and its mean characteristics) hinders the distinction between the endogenous and exogenous effects. In a laboratory experiment that attempts to overcome these difficulties, they find fairness effects but no conformity effects.

## 4.4 Other individual characteristics

Individuals' income, education, age, and gender have been found to correlate with evasion, though the direction and magnitudes of the connections vary (Slemrod, 2007; Blumenthal et al., 2001; Richardson, 2006; Coricelli et al., 2010). Perhaps the one relatively more robust finding is that women are more prone (or at least, not less prone) to comply than men. Theoretically, however, it is not clear how these variables should correlate with evasion, and some may be indirectly capturing other influences such as moral views or peer effects (for instance, moral views or social networks transmitted through education). Income, of course, can play a role in the optimal (rational) level of tax evasion, but predictions vary depending on the modeling assumptions used. In our data, neither age nor gender (rows 24 and 26) is robustly correlated with evasion. Instead, wealthier, more educated individuals, those who own land, and those with higher levels of expenditures are less likely to evade (rows 1, 10, 3, and 15, respectively), and those suffering a negative shock are more likely to avoid paying tax (row 14).<sup>26</sup> This suggests that paying taxes is a normal good that is consumed more by the relatively well off.

## 5 Conclusions

We measure tax evasion, a critical behavior for democracies, in the context of a large-scale and detailed household survey. Tax evasion lies at the core of the relationship between citizens and the state, yet it is difficult to empirically analyze it. Coarse, indirect measures are often imprecise and/or not disaggregated enough to study the underlying behavioral motivations, while survey measures must deal with biases in stated responses. Using list experiments, we directly tackled the bias problem and estimated the incidence of VAT evasion providing direct evidence that social desirability in respondents' claims does not contaminate our measures. Moreover, we examined the resulting incidence and main correlates of tax evasion, reviewing the literature to guide the analysis.

A crucial conclusion is that the theoretical and empirical literature on tax evasion is underdeveloped, especially concerning indirect taxes in general and VAT in particular. This is a severe limitation, since many countries rely heavily on revenue from indirect taxes on goods

---

<sup>26</sup> The standard model of tax evasion views avoidance essentially as a gamble, so risk aversion (which may correlate with income) is potentially important. Since we lack good measures of risk aversion in our data, we leave this aside. Moreover, one view is that the levels of risk aversion required for there to be compliance in equilibrium in the baseline model are excessive Allingham and Sandmo (1972), so other influences must be introduced to explain the data. For example, Bernasconi (1998) extends the baseline model, relaxing the differentiability of the preference function at low levels of risk, thus creating more risk-averse behavior and improving the model's fit with the data.

and services. Also, available (and typically indirect, macro) estimates suggest that indirect tax evasion is important in practice. As noted, these aggregate measures preclude evaluating individual behavioral responses, and this may help account for the relative stagnation of the theory and evidence on indirect tax evasion.

Our rich micro-level data, together with direct evidence that individuals' responses are not contaminated by reporting biases, avoids these problems and is therefore more informative of the incidence and correlates of evasion, and pertinent to develop and test new theories. Also, this is the first list experiment applied in the area of tax evasion, yet the method can be implemented elsewhere and extended to study other forms of evasion beyond the VAT. Combined with direct questions on a random sample of respondents as in our design, this also creates measures of the extent of social desirability bias. Presumably, one key step in consolidating the legitimacy of the state and its enforcement ability in tax contributions is that citizens understand that taxes *ought* to be paid. The absence of social desirability bias in our sample suggests that this norm is not embedded in Colombian society. Examining whether this is also the case in other societies with prevalent VAT (or other forms of tax) evasion is therefore also relevant. List experiments of the sort we described in this paper provide a direct opportunity to do this.

One possible question concerns the external validity of the analyses stemming from the examination of these data. However, VAT evasion is prevalent in many countries and our data, though drawn from Colombia, is sufficiently fine-grained that a careful examination of the mechanisms involved and forces at play is possible, with lessons that can travel beyond the Colombian case.



## References

- Acemoglu, D. (2005). Politics and economics in weak and strong states. *Journal of Monetary Economics*, 52(7), 1199-1226.
- Acemoglu, D., Garcia-Jimeno, C., & Robinson, J. A. (2015). State capacity and economic development: a network approach. *American Economic Review*, 105(8), 2364-2409.
- Allingham, M., & Sandmo, A. (1972). Income tax evasion: a theoretical analysis. *Journal of Public Economics*, 1(3-4), 323-338.
- Alm, J., Jackson, B., & McKee, M. (1992). Estimating the determinants of taxpayer compliance with experimental data. *National Tax Journal*, 52(1), 107-114.
- Alt, J. (1983). The evolution of tax structures. *Public Choice*, 41(1), 181-222.
- Andreoni, J., Erard, B., & Feinstein, J. (1998). Tax compliance. *Journal of economic literature*, 36(2), 818-860.
- Arias, R. (2005). A note on indirect tax evasion. *Asociación Argentina de Economía Política*.
- Avila, J., & Cruz, A. (2007). Recaudar no basta, los impuestos como instrumento de desarrollo. *Dirección de Impuestos y Aduanas Nacionales de Colombia*.
- Bernal, R., Cadena, X., Camacho, A., Cárdenas, J., Fergusson, L., & Ibáñez, A. (2014). Encuesta longitudinal de la Universidad de los Andes (ELCA) – 2013. *Documento CEDE 42/2014, Universidad de Los Andes*.
- Bernasconi, M. (1998). Tax evasion and orders of risk aversion. *Journal of Public Economics*, 67(1), 123-134.
- Besley, T., & Persson, T. (2009). The origins of state capacity: property rights, taxation, and politics. *American Economic Review*, 99(4), 1218-44.
- Besley, T., & Persson, T. (2014). Why do developing countries tax so little? *The Journal of Economic Perspectives*, 28(4), 99-120.
- Blair, G., & Imai, K. (2012). Statistical analysis of list experiments. *Political Analysis*, 20(1), 47-77.
- Blumenthal, M., Christian, C., Slemrod, J., & Smith, M. (2001). Do normative appeals affect tax compliance? Evidence from a controlled experiment in Minnesota. *National Tax Journal*, 54(1), 125-138.
- Bott, K., Cappelen, A., Sørensen, E., & Tungodden, B. (2014). You've got mail: a randomised field experiment on tax evasion. *Discussion paper 26/2014, Department of Economics, Norwegian School of Economics*.
- Christie, E., & Holzner, M. (2006). *What explains tax evasion? An empirical assessment based on European data* (Tech. Rep.). Vienna: The Vienna Institute for International Economic Studies.

- Comisión de expertos para la equidad y la competitividad tributaria. (2015). Informe final presentado al Ministro de Hacienda y Crédito Público. *Presidente: Santiago Rojas, Miembros: Ricardo Bonilla, Rosario Córdoba, Alfredo Lewin, Oscar Darío Morales, Soraya Montoya, Guillermo Perry, Julio Roberto Piza, Miguel Urrutia, Leonardo Villar.*
- Corbacho, A., Fretes, V., & Lora, E. (2013). Recaudar no basta: los impuestos como instrumento de desarrollo. *Inter-American Development Bank.*
- Coricelli, G., Joffily, M., Montmarquette, C., & Villeval, M. (2010). Cheating, emotions, and rationality: an experiment on tax evasion. *Experimental Economics*, 13(2), 226-247.
- Cremer, H., & Gahvari, F. (1993). Tax evasion and optimal commodity taxation. *Journal of Public Economics*, 50(2), 261-275.
- De Jantscher, C. (1986). Problems of administering a value-added tax in developing countries. *IMF Working Paper*, 15/1986.
- Erard, B., & Feinstein, J. (1994). The role of moral sentiments and audit perceptions in tax compliance. *Public Finance/Finances Publique*, 49(1), 70-89.
- Fergusson, L., Molina, C., & Robinson, J. (2017). *The vicious cycle of clientelism and state weakness.* (Unpublished manuscript)
- Fergusson, L., Molina, C. A., & Riaño, J. F. (2017). I sell my vote, and so what? a new database and evidence from Colombia. *Documento CEDE 20/2017, Universidad de Los Andes.*
- Fiorio, C., & D'Amuri, F. (2006). Tax evasion in Italy: an analysis using a tax-benefit microsimulation model. *The IUP Journal of Public Finance*, 4(2), 19-37.
- Fortin, B., Lacroix, G., & Villeval, M. (2007). Tax evasion and social interactions. *Journal of Public Economics*, 91(11), 2089-2112.
- Gemmell, N., & Hasseldine, J. (2012). The tax gap: a methodological review. *Advances in Taxation*, 20(1), 203-231.
- Gómez-Sabaini, J., & Jiménez, J. (2011). Estructura tributaria y evasión impositiva en América Latina. *Working paper, Development Bank of América Latina.*
- Gonzalez-Ocantos, E., de Jonge, C., Meléndez, C., Osorio, J., & Nickerson, D. (2012). Vote buying and social desirability bias: experimental evidence from Nicaragua. *American Journal of Political Science*, 56(1), 202-217.
- Gordon, J. (1989). Individual morality and reputation costs as deterrents to tax evasion. *European Economic Review*, 33(4), 797-805.
- Gordon, J. (1990). Evading taxes by selling for cash. *Oxford Economic Papers*, 42(1), 244-255.
- Grasmick, H., & Bursik Jr, R. (1990). Conscience, significant others, and rational choice:

- extending the deterrence model. *Law and society review*, 24(3), 837-862.
- Hanlon, M., & Heitzman, S. (2010). A review of tax research. *Journal of Accounting and Economics*, 50(2), 127-178.
- Imai, K. (2011). Multivariate regression analysis for the item count technique. *American Statistical Association*, 106(494), 407-416.
- Katz, B., & Owen, J. (2013). Exploring tax evasion in the context of political uncertainty. *Economic Systems*, 37(2), 141-154.
- Khlif, H., & Achek, I. (2015). The determinants of tax evasion: a literature review. *International Journal of Law and Management*, 57(5), 486-497.
- Kirchler, E. (2007). *The economic psychology of tax behaviour*. Cambridge: Cambridge University Press.
- Leamer, E. (1985). Sensitivity analyses would help. *American Economic Review*, 75(3), 308-313.
- Litina, A., & Palivos, T. (2016). Corruption, tax evasion and social values. *Journal of Economic Behavior & Organization*, 124(1), 164-177.
- Luttmer, E., & Singhal, M. (2014). Tax morale. *Journal of Economic Perspectives*, 28(4), 149-168.
- Manski, C. (1993). Identification of endogenous social effects: the reflection problem. *The Review of Economic Studies*, 60(3), 531-542.
- Marrelli, M. (1984). On indirect tax evasion. *Journal of Public Economics*, 25(1), 181-196.
- Matthews, K., & Lloyd Williams, J. (2001). The VAT-evading firm and VAT evasion: an empirical analysis. *International Journal of the Economics of Business*, 8(1), 39-49.
- McGee, R. (2011). *The ethics of tax evasion: perspectives in theory and practice*. New York: Springer Science & Business Media.
- Montalvo, J., & Reynal-Querol, M. (2005). Ethnic polarization, potential conflict and civil war. *American Economic Review*, 95(3), 796-816.
- Myles, G., & Naylor, R. (1996). A model of tax evasion with group conformity and social customs. *European Journal of Political Economy*, 12(1), 49-66.
- Pachón, M., & Sánchez, F. (2014). Base de datos sobre resultados electorales. *Documento CEDE 29/2014, Universidad de Los Andes*.
- Parra, O., & Patiño, R. (2010). Evasión de impuestos nacionales en Colombia: años 2001-2009. *Revista Facultad de Ciencias Económicas, Universidad Militar Nueva Granada*, 18(2), 177-198.
- Picur, R., & Riahi-Belkaoui, A. (2006). The impact of bureaucracy, corruption and tax compliance. *Review of Accounting and Finance*, 5(2), 174-180.
- Pomeranz, D. (2015). No taxation without information: deterrence and self-enforcement in

- the value added tax. *American Economic Review*, 105(8), 2539-2569.
- Reynal-Querol, M. (2002). Ethnicity, political systems, and civil wars. *Journal of Conflict Resolution*, 46(1), 29-54.
- Riahi-Belkaoui, A. (2004). Relationship between tax compliance internationally and selected determinants of tax morale. *Journal of International Accounting, Auditing and Taxation*, 13(2), 135-143.
- Richardson, G. (2006). Determinants of tax evasion: a cross-country investigation. *Journal of International Accounting, Auditing and Taxation*, 15(2), 150-169.
- Richardson, G. (2008). The relationship between culture and tax evasion across countries: additional evidence and extensions. *Journal of International Accounting, Auditing and Taxation*, 17(2), 67-78.
- Sala-i Martin, X. (1997). I just ran two million regressions. *American Economic Review*, 87(2), 178-183.
- Slemrod, J. (1998). On voluntary compliance, voluntary taxes, and social capital. *National Tax Journal*, 51(3), 485-491.
- Slemrod, J. (2007). Cheating ourselves: the economics of tax evasion. *The Journal of Economic Perspectives*, 21(1), 25-48.
- Slemrod, J., & Yitzhaki, S. (2002). Tax avoidance, evasion, and administration. In Auerbach & Feldstein (Eds.), *Handbook of Public Economics* (Vol. 3, p. 1423-1470). Elsevier.
- Smith, K. (1992). *Reciprocity and fairness: positive incentives for tax compliance*. Chicago: American Bar Foundation.
- Spicer, M., & Lundstedt, S. (1976). Understanding tax evasion. *Public Finance*, 31(2), 295-305.
- Torgler, B. (2003). Tax morale, rule-governed behaviour and trust. *Constitutional Political Economy*, 14(2), 119-140.
- Virmani, A. (1989). Indirect tax evasion and production efficiency. *Journal of Public Economics*, 39(2), 223-237.
- Wang, L., & Conant, J. (1988). Corporate tax evasion and output decisions of the uncertain monopolist. *National Tax Journal*, 41(4), 579-581.
- Webley, P. (1991). *Tax evasion: an experimental approach*. Cambridge: Cambridge University Press.
- Yaniv, G. (1988). Withholding and non-withheld tax evasion. *Journal of Public Economics*, 35(2), 183-204.
- Yaniv, G. (1995). A note on the tax evading firm. *National Tax Journal*, 41(1), 113-120.
- Zídková, H. (2014). Determinants of VAT gap in EU. *Prague Economic Papers*, 23(4), 514-530.

# A Online appendix (Not for publication)

## A.1 Variable definition and sources

Table A-1: Variable definition and sources

Variable	Description
<i>Age</i>	Respondent’s age in years.
<i>Agree with bribery</i>	Equals 1 if respondent “totally agrees” or “agrees” with the statement: “As things are, sometimes paying a bribe is justified.”
<i>Authorities violate the law</i>	Equals 1 if respondent “totally agrees” or “agrees” with the statement: “To capture criminals, authorities should sometimes violate the law.”
<i>Catholic</i>	Equals 1 for Catholics (self-declared).
<i>Commerce sector</i>	Equals 1 if commerce is one of three sectors generating more formal employment in the municipality (2012). <i>Ministry of Health and Social Protection</i> .
<i>Contributes to social security</i>	Equals 1 if respondent pays contributions to the social security system (thus excluding members in the subsidized social security system).
<i>Dark</i>	Respondent’s skin color based on color palette (assessed by interviewer). The palette is numbered from 1 to 11 (1 = lightest color). <i>Dark</i> Equals 1 for colors greater than or equal to 5.
<i>Education</i>	Educational attainment. Equals 1 if respondent has some secondary education or more.
<i>Employed household head</i>	Equals 1 if household head was employed in the week preceding the survey.
<i>Employment</i>	Equal 1 if respondent was employed the week previous to the survey.
<i>Employment firms 2+</i>	Total formal employment in firms with two or more employees as a proportion of total formal employment (2012). <i>Ministry of Health and Social Protection</i> .
<i>Evangelical/Pentecostal</i>	Equals 1 for Evangelical/Pentecostal (self-declared).
<i>FEA</i>	Equals 1 if household is a <i>Familias en acción</i> (main conditional cash transfer for the poor with school-age children) beneficiary.
<i>Formal credit</i>	Equals 1 if household has any formal credit.
<i>Fractionalization</i>	$F_j = 1 - \sum_{i=1}^N \pi_{ij}^2$ , where $\pi_{ij}$ is the vote share for the mayoral candidate (in 2011) $i$ in municipality $j$ . See <a href="#">Montalvo and Reynal-Querol (2005)</a> . <a href="#">Pachón and Sánchez (2014)</a> .
<i>Frequency of voting</i>	Equals 1 if respondent “Always votes in elections” or “Votes in most elections”, 0 otherwise (“Rarely votes in elections” or “Never voted”).
<i>Get help</i>	Equals 1 if household answers “yes” to at least one of the following questions: “During the past 12 months, do any members of the household receive money or in-kind aid...” a. “...from relatives or friends living in Colombia?”, b. “from relatives or friends living abroad?” c. “for alimony?”, d. “from international organizations (WFP, UNICEF, ICRC)?”, e. “from NGOs?”, f. “from the church or other religious organizations?”, g. “from other persons, entities or organizations?”.

Continued on next page

---



---

**Table A-1 – Variable definition and sources, continued from previous page**

---



---

<b>Variable</b>	<b>Description</b>
<i>Gov. against inequality</i>	Equals 1 if respondent “totally agrees” or “agrees” with the statement: “The government should implement strong policies to reduce inequality between rich and poor.”
<i>Government role</i>	Equals 1 if respondent “totally agrees” or “agrees” with the statement: “Government is primarily responsible for ensuring the welfare of the people.”
<i>Guerrillas</i>	Number of violent events per year perpetrated by guerrillas per 100,000 inhabitants (average 2010–2012). <i>Conflict analysis resource center (CERAC) and Universidad del Rosario.</i>
<i>HH expenses</i>	Per capita household expenses (Colombian pesos). See <a href="#">Bernal et al. (2014)</a> .
<i>HH food expenses</i>	Household expenses in food (Colombian pesos). See <a href="#">Bernal et al. (2014)</a> .
<i>Homeowner</i>	Equals 1 if the household residence is “own, fully paid” or “own, being paid”. Equals 0 otherwise (“rented” or “in usufruct or other type of tenure”).
<i>Homicide rate</i>	Homicide rate by 100,000 inhabitants (average 2010–2012). <i>Medicina Legal.</i>
<i>Household with spouse</i>	Equals 1 if household is inhabited by household head and spouse.
<i>Independent</i>	Equals 1 if working independently is the most important job during the previous month.
<i>Justice into own hands</i>	Equals 1 if respondent “totally agrees” or “agrees” with the statement: “When the government does not punish criminals, it is okay that people take justice into their own hands.”
<i>Lands</i>	Equals 1 if respondent reports owning land.
<i>Male household head</i>	Equals 1 if household head is male.
<i>Neighbor cell phones</i>	Equals 1 if person has the cell phone number of at least half of her neighbors.
<i>Neighbor loans</i>	Equals 1 if a person thinks that at least half of her neighbors would lend her money.
<i>Negative reciprocity</i>	Equals 1 if respondent “totally agrees” or “agrees” with the statement: “Whoever hurts me, pays for it.”
<i>No debts</i>	Equals 1 if household has not any type of debt. Equals 0 otherwise (“debt without credit”, “formal credit”, “informal credit” or “formal-informal credit”).
<i>No sewage</i>	Equals 1 if household dwelling has no sewage system.
<i>Not in organization</i>	Equals 1 if respondent does not belong to any organization (options included are <i>Juntas de acción comunal</i> , charity organization, community organization, religious organization, organizations supported or promoted by the state, ethnic organization, educational organization, labor union, cooperative of work or union of producers, organization of environment conservation, cultural or sports organization, other).
<i>Nuclear family</i>	Equals 1 household is comprised of: household head and spouse, with or without children; or, household head without spouse but with children).
<i>Other religion</i>	Equals 1 for believers of religions other than Catholic, Evangelical, or Pentecostal (self-declared).
<i>Overcrowded</i>	Equals 1 if ratio of number of residents to number of bedrooms is greater than three in rural households, or greater than or equal to three in urban households.

Continued on next page

---



---

**Table A-1 – Variable definition and sources, continued from previous page**

---



---

<b>Variable</b>	<b>Description</b>
<i>Own welfare</i>	Equals 1 if respondent “totally agrees” or “agrees” with the statement: “Each individual is responsible for their own welfare.”
<i>Paramilitaries</i>	Number of violent events per year perpetrated by paramilitaries per 100,000 inhabitants (average 2010–2012). <i>Conflict analysis resource center (CERAC) and Universidad del Rosario</i> .
<i>People in household</i>	Number of household residents.
<i>Polarization</i>	Reynal-Querol (2002) polarization index. $P_j = 1 - \sum_{i=1}^N \pi_{ij} \left( \frac{1/2 - \pi_{ij}}{1/2} \right)^2$ , where $\pi_{ij}$ is the vote share for the mayoral candidate (in 2011) $i$ in municipality $j$ . <i>Pachón and Sánchez (2014)</i> .
<i>Pop. density</i>	Population divided by total area ( $km^2$ ) in the municipality.
<i>Popular vote</i>	Equals 1 if respondent “totally agrees” or “agrees” with the statement: “It is important that rulers are elected by popular vote.”
<i>Positive reciprocity</i>	Equals 1 if respondent “totally agrees” or “agrees” with the statement: “You always have to help those who help you.”
<i>Regions</i>	Regions included in fixed effects. Urban regions in the survey are: <i>Atlántica, Oriental, Central, Pacífica, Bogotá</i> . Rural regions include: <i>Atlántica-Media, Cundi-Boyacense, Eje Cafetero, Centro-Oriente</i> .
<i>Rural population</i>	Proportion of rural population in the municipality (average 2006–2008). <i>DANE</i> .
<i>Savings</i>	Equals 1 if respondent answers “yes” to: “Do you usually save some of the income you receive?”.
<i>Secret ballot</i>	Equals 1 if respondent answers “yes” to: “Do you think that the ballot is secret?”
<i>Send help</i>	Equals 1 if household answers “yes” to at least one of the following questions: “During the past 12 months, did any members of the household send money or in-kind aid...” a. “to relatives or friends who live in Colombia?”, b. “to relatives or friends who live abroad?”, c. “for alimony?” d. “to other persons, entities or organizations?”.
<i>Shock</i>	Equals 1 if household reports any major destabilizing negative event during the previous three years.
<i>Social program beneficiary</i>	Equals 1 if household benefits from any of the following programs: <i>Familias en acción</i> (main conditional cash transfer for the poor with school-age children), programs for the elderly, <i>SENA</i> training programs, <i>Red Juntos - Unidos</i> (program that provides social services to displaced families with the lowest levels of poverty), <i>ICBF</i> programs for children, aid for displaced people, support to households affected by natural disasters, or “other programs”.
<i>Social security</i>	Equals 1 if respondent is affiliated to social security.
<i>State presence</i>	Raw total of local state agencies, local municipality employees, and national-level municipality employees (per capita in 1995). <i>Acemoglu, Garcia-Jimeno, and Robinson (2015)</i> .
<i>Stratum 1, 2</i>	Socio-economic stratum, based on classification of household residence (used to target utility subsidies).

Continued on next page

---



---

**Table A-1 – Variable definition and sources, continued from previous page**

---



---

<b>Variable</b>	<b>Description</b>
<i>Use of violence</i>	Equals 1 if respondent “totally agrees” or “agrees” with the statement: “Sometimes the use of violence is justified.”
<i>Vote for the same party</i>	Equals 1 if respondent “Always votes for the same party” or “Almost always votes for the same party”. Equals 0 otherwise (“Votes for different parties” or “Always votes blank”).
<i>Wealth</i>	First principal component following a <i>principal component analysis</i> on a set of reported household assets and dwelling characteristics. See <a href="#">Bernal et al. (2014)</a> .
<i>Win margin</i>	Difference between the vote shares of the winner and runner-up in the 2011 mayoral election. <a href="#">Pachón and Sánchez (2014)</a> .
<i>Woman</i>	Equals 1 if respondent is female.
<i>Workers per firm</i>	Average number of formal workers per firms by municipality (2012). <i>Ministry of Health and Social Protection</i> .

---

**Notes:** Source is Elca 2013 unless otherwise stated at the end of each description.



## A.2 Balance between treatment and control groups in list experiments

We corroborate that respondents assigned to treatment and control lists, and the direct question, have similar observable characteristics. For a set of observables  $\mathbf{X}$ , we check both the bivariate relationship between group assignment and observables:

$$\Pr(T_i = m) = f(x_i\beta_m) \quad \text{with } x_i \in \mathbf{x},$$

and the multivariate regression,

$$\Pr(T_i = m) = f(\mathbf{x}'\beta_m),$$

where  $m$  represents each group (*Treatment*, *Control 1*, and *Control 2*). We estimate the marginal effects of multinomial probit models.

Since randomization was stratified at the regional level, in both types of regressions we include region fixed effects. We also estimated separate regressions for each region, with similar results, but present only these aggregate results to save space. Similarly, we also estimated simple probit and linear probability models for dichotomous indicators of each treatment condition as the dependent variable, and again found no systematic evidence of imbalance.

Table [A-2](#) shows balance using observables in 2010 and Table [A-3](#) in 2013 for the tax evasion experiment.

**Table A-2: Balance on covariates at the baseline (2010): tax evasion list experiment**

Variables	Urban sample						Rural sample					
	Bivariate			Multivariate			Bivariate			Multivariate		
	<i>Treatment</i>	<i>Control 1</i>	<i>Control 2</i>	<i>Treatment</i>	<i>Control 1</i>	<i>Control 2</i>	<i>Treatment</i>	<i>Control 1</i>	<i>Control 2</i>	<i>Treatment</i>	<i>Control 1</i>	<i>Control 2</i>
Age	-0.001 (0.00)	0.001 (0.00)	-0.000 (0.00)	-0.001 (0.00)	0.001 (0.00)	-0.000 (0.00)	-0.001 (0.00)	0.000 (0.00)	0.001 (0.00)	-0.001* (0.00)	0.001 (0.00)	0.001 (0.00)
Male household head	-0.001 (0.01)	-0.000 (0.01)	0.001 (0.01)	0.008 (0.02)	-0.022 (0.02)	0.014 (0.02)	0.002 (0.02)	-0.023 (0.02)	0.022 (0.02)	-0.004 (0.03)	-0.006 (0.03)	0.009 (0.03)
Education	0.001 (0.00)	-0.002 (0.00)	0.001 (0.00)	0.001 (0.00)	-0.000 (0.00)	-0.000 (0.00)	-0.001 (0.00)	-0.000 (0.00)	0.001 (0.00)	-0.001 (0.00)	-0.000 (0.00)	0.002 (0.00)
Employed household head	0.005 (0.02)	0.003 (0.02)	-0.007 (0.02)	-0.002 (0.02)	0.011 (0.02)	-0.009 (0.02)	-0.007 (0.02)	0.018 (0.02)	-0.011 (0.02)	-0.015 (0.02)	0.027 (0.02)	-0.012 (0.02)
Savings	-0.004 (0.02)	-0.010 (0.02)	0.014 (0.02)	-0.009 (0.02)	-0.007 (0.02)	0.016 (0.02)	-0.018 (0.03)	-0.010 (0.03)	0.028 (0.03)	-0.018 (0.03)	-0.009 (0.03)	0.027 (0.03)
Not in organization	0.000 (0.00)	-0.000 (0.00)	-0.000 (0.00)	0.000 (0.00)	0.000 (0.00)	-0.000 (0.00)	0.001 (0.00)	-0.000 (0.00)	-0.000 (0.00)	0.001 (0.00)	-0.000 (0.00)	-0.000 (0.00)
Social security	0.024 (0.02)	0.001 (0.02)	-0.026 (0.02)	0.030 (0.02)	-0.006 (0.02)	-0.024 (0.02)	-0.007 (0.03)	-0.001 (0.03)	0.008 (0.03)	0.004 (0.03)	-0.003 (0.03)	-0.002 (0.03)
Contributes to social security	-0.006 (0.01)	-0.000 (0.01)	0.006 (0.01)	-0.015 (0.02)	0.011 (0.02)	0.004 (0.02)	-0.038 (0.03)	0.013 (0.03)	0.026 (0.02)	-0.031 (0.04)	0.016 (0.03)	0.015 (0.03)
Household with spouse	0.006 (0.01)	-0.015 (0.01)	0.009 (0.01)	0.006 (0.02)	-0.021 (0.02)	0.015 (0.02)	0.000 (0.02)	0.028 (0.02)	-0.028 (0.02)	-0.004 (0.03)	0.026 (0.03)	-0.022 (0.03)
Wealth	0.003 (0.01)	-0.003 (0.01)	-0.000 (0.01)	0.007 (0.01)	0.004 (0.01)	-0.011 (0.01)	-0.010 (0.01)	0.008 (0.01)	0.002 (0.01)	-0.008 (0.01)	0.014 (0.01)	-0.006 (0.01)
People in household	0.001 (0.01)	0.002 (0.00)	-0.004 (0.00)	0.003 (0.01)	-0.000 (0.00)	-0.003 (0.01)	0.003 (0.00)	-0.005 (0.00)	0.002 (0.00)	-0.001 (0.00)	-0.003 (0.01)	0.004 (0.01)
Overcrowded	0.001 (0.02)	-0.019 (0.02)	0.018 (0.02)	0.004 (0.02)	-0.019 (0.02)	0.016 (0.02)	-0.021 (0.02)	0.006 (0.02)	0.015 (0.02)	-0.013 (0.02)	-0.008 (0.02)	0.022 (0.02)
Homeowner	-0.002 (0.01)	-0.004 (0.01)	0.007 (0.01)	0.005 (0.02)	-0.013 (0.01)	0.008 (0.02)	-0.005 (0.02)	-0.013 (0.01)	0.018 (0.01)	0.006 (0.02)	-0.016 (0.02)	0.010 (0.02)
No debts	-0.018 (0.01)	0.024* (0.01)	-0.006 (0.01)	-0.020 (0.02)	0.030* (0.02)	-0.011 (0.02)	-0.018 (0.02)	0.006 (0.01)	0.012 (0.02)	-0.011 (0.02)	0.017 (0.02)	-0.006 (0.02)
Shock	0.011 (0.02)	0.011 (0.01)	-0.022 (0.02)	0.009 (0.02)	0.015 (0.01)	-0.025 (0.02)	-0.021 (0.02)	0.013 (0.02)	0.008 (0.02)	-0.023 (0.02)	0.011 (0.02)	0.012 (0.02)
HH expenses (log)	-0.002 (0.01)	-0.004 (0.01)	0.007 (0.01)	0.005 (0.02)	-0.013 (0.02)	0.007 (0.02)	0.014 (0.01)	-0.020* (0.01)	0.006 (0.01)	0.002 (0.02)	-0.014 (0.02)	0.012 (0.02)
HH food expenses	-0.010 (0.01)	0.003 (0.01)	0.008 (0.01)	-0.022 (0.02)	0.013 (0.02)	0.009 (0.02)	0.022* (0.01)	-0.019 (0.01)	-0.003 (0.01)	0.030 (0.02)	-0.006 (0.02)	-0.024 (0.02)
Nuclear family	-0.005 (0.02)	0.002 (0.01)	0.002 (0.02)	-0.002 (0.02)	-0.001 (0.02)	0.002 (0.02)	-0.002 (0.02)	-0.000 (0.02)	0.003 (0.01)	-0.006 (0.02)	-0.000 (0.02)	0.007 (0.02)
Formal credit	-0.008 (0.01)	0.011 (0.01)	-0.003 (0.01)	0.001 (0.02)	-0.005 (0.02)	0.003 (0.02)	-0.027 (0.02)	-0.005 (0.02)	0.032* (0.02)	-0.027 (0.02)	-0.009 (0.03)	0.036* (0.02)
Get help	0.008 (0.02)	-0.004 (0.01)	-0.005 (0.02)	0.010 (0.02)	-0.007 (0.02)	-0.003 (0.02)	0.017 (0.02)	-0.016 (0.02)	-0.001 (0.02)	0.016 (0.02)	-0.010 (0.02)	-0.006 (0.02)
Send help	0.005 (0.02)	-0.013 (0.02)	0.009 (0.02)	0.002 (0.02)	-0.016 (0.02)	0.014 (0.02)	-0.045 (0.03)	0.029 (0.03)	0.017 (0.03)	-0.052* (0.03)	0.030 (0.03)	0.022 (0.03)
Social program beneficiary	0.016 (0.02)	0.008 (0.01)	-0.024 (0.01)	0.028 (0.03)	0.018 (0.02)	-0.046** (0.02)	0.010 (0.02)	-0.005 (0.02)	-0.005 (0.02)	-0.004 (0.03)	0.020 (0.03)	-0.016 (0.03)
FEA	0.008 (0.02)	0.004 (0.02)	-0.012 (0.02)	-0.012 (0.04)	-0.022 (0.03)	0.034 (0.03)	0.013 (0.02)	-0.014 (0.02)	0.001 (0.02)	0.011 (0.03)	-0.030 (0.03)	0.019 (0.03)
Stratum 1	-0.016 (0.02)	0.011 (0.02)	0.005 (0.02)	-0.035 (0.02)	0.035 (0.02)	-0.000 (0.02)						
Stratum 2	-0.003 (0.02)	0.021 (0.01)	-0.018 (0.02)	-0.018 (0.02)	0.037* (0.02)	-0.019 (0.02)						
Wealth (rural): quintile 1							0.007 (0.02)	-0.013 (0.02)	0.006 (0.02)	0.007 (0.04)	-0.004 (0.04)	-0.002 (0.04)
Wealth (rural): quintile 2							0.013 (0.02)	0.022 (0.02)	-0.035* (0.02)	0.015 (0.03)	0.027 (0.04)	-0.042 (0.04)
Wealth (rural): quintile 3							0.021 (0.02)	-0.037* (0.02)	0.017 (0.02)	0.016 (0.03)	-0.017 (0.03)	0.001 (0.03)
Wealth (rural): quintile 4							-0.017 (0.02)	0.021 (0.02)	-0.004 (0.02)	-0.010 (0.03)	0.024 (0.03)	-0.015 (0.03)
No sewage							0.006 (0.02)	-0.023 (0.02)	0.018 (0.02)	0.010 (0.02)	-0.026 (0.02)	0.015 (0.02)

**Notes:** The table reports marginal effects from multinomial probit models, with standard errors in parentheses. All regressions include region fixed effects. *Treatment* refers to respondents who were presented with the list that included a sensitive item, *Control 1* received the list without the sensitive item followed by the direct question, and *Control 2* was asked the direct question. For variable definitions, see Appendix Table A-1. \* is significant at the 10% level, \*\* is significant at the 5% level, \*\*\* is significant at the 1% level.

**Table A-3: Balance on covariates at the follow-up (2013): tax evasion list experiment**

Variables	Urban sample						Rural sample					
	Bivariate			Multivariate			Bivariate			Multivariate		
	Treatment	Control 1	Control 2	Treatment	Control 1	Control 2	Treatment	Control 1	Control 2	Treatment	Control 1	Control 2
Age	-0.000 (0.00)	0.001 (0.00)	-0.000 (0.00)	-0.000 (0.00)	0.001* (0.00)	-0.001 (0.00)	-0.001 (0.00)	0.000 (0.00)	0.001 (0.00)	-0.001 (0.00)	0.000 (0.00)	0.001 (0.00)
Male household head	-0.001 (0.01)	-0.007 (0.01)	0.008 (0.01)	-0.007 (0.02)	-0.012 (0.02)	0.020 (0.02)	-0.001 (0.02)	-0.011 (0.02)	0.012 (0.02)	0.004 (0.03)	0.005 (0.02)	-0.009 (0.02)
Education	-0.000 (0.00)	0.000 (0.00)	0.000 (0.00)	-0.001 (0.00)	0.002 (0.00)	-0.001 (0.00)	0.001 (0.00)	-0.002 (0.00)	0.001 (0.00)	0.001 (0.00)	-0.003 (0.00)	0.002 (0.00)
Employed household head	0.011 (0.02)	-0.018 (0.02)	0.007 (0.02)	0.012 (0.02)	-0.014 (0.02)	0.002 (0.02)	0.015 (0.02)	0.005 (0.02)	-0.020 (0.02)	0.004 (0.02)	0.014 (0.02)	-0.018 (0.02)
Savings	0.004 (0.02)	-0.018 (0.02)	0.014 (0.02)	0.006 (0.02)	-0.015 (0.02)	0.008 (0.02)	0.029 (0.02)	0.019 (0.02)	-0.048*** (0.02)	0.029 (0.02)	0.017 (0.02)	-0.046*** (0.02)
Not in organization	0.000 (0.02)	-0.020 (0.02)	0.020 (0.01)	-0.010 (0.02)	-0.017 (0.02)	0.027* (0.02)	-0.011 (0.02)	-0.016 (0.01)	0.026* (0.01)	-0.005 (0.02)	-0.020 (0.02)	0.025 (0.02)
Social security	0.025 (0.03)	-0.038 (0.02)	0.013 (0.03)	0.034 (0.03)	-0.047* (0.03)	0.013 (0.03)	0.046 (0.03)	-0.034 (0.03)	-0.012 (0.04)	0.042 (0.04)	-0.049 (0.04)	0.007 (0.04)
Contributes to social security	0.012 (0.01)	0.005 (0.01)	-0.016 (0.01)	0.031* (0.02)	0.004 (0.02)	-0.035** (0.02)	-0.024 (0.02)	0.006 (0.02)	0.018 (0.02)	-0.004 (0.02)	-0.015 (0.02)	0.019 (0.02)
Household with spouse	-0.003 (0.01)	-0.007 (0.01)	0.010 (0.01)	-0.003 (0.02)	-0.021 (0.02)	0.024 (0.02)	0.010 (0.02)	0.023 (0.02)	-0.033* (0.03)	0.004 (0.02)	0.045* (0.02)	-0.049** (0.02)
Wealth	-0.005 (0.01)	0.002 (0.01)	0.003 (0.01)	-0.003 (0.01)	0.006 (0.01)	-0.003 (0.01)	-0.014* (0.01)	0.011 (0.01)	0.003 (0.01)	-0.024 (0.02)	-0.014 (0.02)	0.038** (0.02)
People in household	0.005 (0.01)	0.001 (0.00)	-0.006 (0.01)	0.007 (0.01)	-0.003 (0.00)	-0.005 (0.01)	-0.002 (0.00)	-0.002 (0.00)	0.004 (0.00)	-0.004 (0.01)	0.002 (0.00)	0.002 (0.01)
Overcrowded	-0.038** (0.02)	0.004 (0.02)	0.034** (0.02)	-0.035* (0.02)	0.008 (0.02)	0.027 (0.02)	-0.003 (0.02)	0.039* (0.02)	-0.036 (0.02)	0.000 (0.03)	0.039* (0.02)	-0.039 (0.03)
Homeowner	-0.007 (0.01)	-0.008 (0.01)	0.015 (0.01)	-0.005 (0.01)	-0.020 (0.01)	0.025* (0.01)	0.016 (0.01)	-0.016 (0.02)	-0.000 (0.01)	0.026* (0.02)	-0.023 (0.02)	-0.004 (0.02)
No debts	0.016 (0.02)	-0.002 (0.02)	-0.013 (0.01)	0.050 (0.03)	-0.068** (0.03)	0.018 (0.03)	0.014 (0.01)	-0.011 (0.02)	-0.003 (0.01)	0.039 (0.03)	-0.014 (0.03)	-0.025 (0.03)
Shock	-0.019 (0.01)	-0.001 (0.02)	0.020 (0.02)	-0.023 (0.01)	-0.002 (0.02)	0.025* (0.01)	0.009 (0.02)	-0.007 (0.02)	-0.002 (0.02)	0.006 (0.02)	-0.006 (0.02)	-0.000 (0.02)
HH expenses (log)	-0.010 (0.01)	-0.001 (0.01)	0.011 (0.01)	-0.016 (0.03)	-0.021 (0.02)	0.037 (0.02)	-0.010 (0.01)	-0.001 (0.01)	0.012 (0.01)	-0.011 (0.02)	-0.019 (0.02)	0.030 (0.02)
HH food expenses (log)	-0.010 (0.01)	0.012 (0.01)	-0.002 (0.01)	-0.011 (0.02)	0.029* (0.02)	-0.018 (0.02)	-0.004 (0.01)	0.005 (0.01)	-0.001 (0.01)	0.010 (0.02)	0.022 (0.02)	-0.033 (0.02)
Nuclear family	-0.009 (0.01)	0.010 (0.02)	-0.001 (0.02)	-0.005 (0.02)	0.001 (0.02)	0.004 (0.02)	-0.004 (0.01)	0.018 (0.01)	-0.013 (0.01)	-0.009 (0.02)	0.033* (0.02)	-0.024 (0.02)
Formal credit	0.003 (0.02)	0.018 (0.02)	-0.021 (0.01)	-0.031 (0.03)	0.075*** (0.03)	-0.044* (0.02)	0.001 (0.01)	-0.008 (0.01)	0.007 (0.01)	-0.028 (0.03)	0.003 (0.03)	0.025 (0.03)
Get help	0.008 (0.01)	-0.025* (0.01)	0.017 (0.01)	0.010 (0.02)	-0.028* (0.01)	0.018 (0.01)	0.005 (0.01)	-0.031** (0.01)	0.026* (0.02)	0.005 (0.02)	-0.026* (0.02)	0.022 (0.02)
Send help	0.019 (0.02)	0.007 (0.02)	-0.026* (0.01)	0.012 (0.02)	0.010 (0.02)	-0.023 (0.02)	-0.001 (0.02)	-0.008 (0.02)	0.009 (0.02)	-0.005 (0.02)	-0.001 (0.02)	0.006 (0.02)
Social program beneficiary	-0.006 (0.01)	0.018 (0.01)	-0.012 (0.01)	0.004 (0.02)	0.005 (0.02)	-0.009 (0.02)	0.009 (0.01)	0.002 (0.02)	-0.011 (0.02)	-0.018 (0.02)	0.015 (0.02)	0.003 (0.02)
FEA	-0.013 (0.02)	0.026 (0.02)	-0.013 (0.02)	-0.034 (0.02)	0.031 (0.02)	0.003 (0.02)	0.022 (0.02)	-0.007 (0.02)	-0.015 (0.02)	0.032 (0.02)	-0.012 (0.02)	-0.020 (0.02)
Stratum 1	0.002 (0.01)	0.006 (0.02)	-0.008 (0.01)	-0.018 (0.03)	0.035 (0.02)	-0.017 (0.02)						
Stratum 2	-0.014 (0.01)	0.023* (0.01)	-0.009 (0.01)	-0.030 (0.02)	0.043** (0.02)	-0.014 (0.02)						
Wealth (rural): quintile 1							0.009 (0.02)	0.008 (0.02)	-0.017 (0.02)	-0.035 (0.05)	-0.090* (0.05)	0.125** (0.05)
Wealth (rural): quintile 2							0.020 (0.02)	-0.025 (0.02)	0.005 (0.02)	-0.015 (0.04)	-0.102*** (0.04)	0.117*** (0.04)
Wealth (rural): quintile 3							0.015 (0.02)	-0.038** (0.02)	0.023 (0.03)	-0.011 (0.03)	-0.105*** (0.03)	0.115*** (0.03)
Wealth (rural): quintile 4							-0.026 (0.02)	0.002 (0.02)	0.024 (0.03)	-0.032 (0.03)	-0.064** (0.03)	0.096*** (0.03)
No sewage							0.002 (0.02)	0.004 (0.02)	-0.005 (0.02)	0.013 (0.02)	-0.003 (0.02)	-0.009 (0.02)

**Notes:** The table reports marginal effects from multinomial probit models, with standard errors in parentheses. All regressions include region fixed effects. *Treatment* refers to respondents who were presented with the list that included a sensitive item, *Control 1* received the list without the sensitive item followed by the direct question, and *Control 2* was asked the direct question. For variable definitions, see Appendix Table A-1. \* is significant at the 10% level, \*\* is significant at the 5% level, \*\*\* is significant at the 1% level.

### A.3 Testing the no design and no liar assumptions in list experiments

In this Appendix we test for the ‘no design’ and ‘no liar’ assumptions in our list experiments, following Blair and Imai (2012). The test for the former compares the predicted average difference in answers to control items under treatment vs. control. With  $Y_i(0)$ ,  $Z_{i,J+1}^*$ ,  $Y_i$  and  $T_i$  specified as above, let  $\pi_{yz} = \Pr(Y_i(0), Z_{i,J+1}^* = z)$  represent the proportion of the population in each type  $(Y_i(0), Z_{i,J+1}^*)$ . If there are no design effects, these proportions can be computed for all  $y = 0, \dots, J$  as follows:

$$\begin{aligned}\pi_{y1} &= \Pr(Y_i \leq y | T_i = 0) - \Pr(Y_i \leq y | T_i = 1), \\ \pi_{y0} &= \Pr(Y_i \leq y | T_i = 1) - \Pr(Y_i \leq y - 1 | T_i = 0).\end{aligned}$$

Proportions  $\pi_{y1}$  and  $\pi_{y0}$  always take positive values. But with design effects, estimated proportions can be negative (for example, see Table 5 in Blair and Imai (2012)). To test for design effects, one can therefore evaluate whether the proportion of the population in each type ( $\pi_{yz}$ ) is jointly nonnegative.<sup>27</sup> Panel A in Table A-4 shows that no single estimated proportion is negative for either experiment, so the test suggests there is no evidence to reject the null hypothesis of no design effects.

To test the ‘no liar’ assumption, we can evaluate the two most common sources of untruthful answers: ceiling and floor effects. These occur when the respondent engages in either none or all of the behaviors, and thus feels exposed if he or she answers truthfully. In Table 1, the bulk of the answers in the treated lists (94.5%) are larger than zero and smaller than the maximum (five) number of items people can list. This reflects that, since the original instrument design, we included option items that are likely to be negatively correlated with each other, as well as at least one very frequent behavior.

We also test for floor and ceiling effects more formally by estimating the model under the no liar assumption, and comparing it to an alternative model allowing for floor and ceiling effects. Based on different information criteria, if the data supports the second model, there is evidence to reject the null of no floor or ceiling effects.<sup>28</sup> Panel B of Table A-4 reports the results. Regardless of the criterion used, Schwarz’s BIC or Akaike’s AIC, the preferred

---

<sup>27</sup> This test, however, has limitations: there can be design effects with positive  $\pi_{y1}$  and  $\pi_{y0}$ . Also, a higher probability of positive answers to the sensitive item reduces the likelihood of rejecting the null of no design effects.

<sup>28</sup> Since the model is identified under the no floor or ceiling effects assumption, we must make additional assumptions to estimate the alternative, allowing for these effects. To do so, we follow Blair and Imai (2012) and consider that respondents’ truthful answers to the sensitive item are independent of their answers for control items, conditional upon the pretreatment covariates.

model includes no floor or ceiling effects, so this test fails to reject the null of no floor or ceiling effects. Furthermore, these results hold either with covariates (Columns 1 and 2) or when the basic set of covariates in Table 2 are included.<sup>29</sup>

**Table A-4: Testing assumptions in the list experiments**

	(1)	(2)	(3)	(4)
<b>Panel A: No design effects</b>				
Response value ( $y$ )	Estimated proportions with response $y$ to control items and...			
	...not following sensitive behavior ( $\hat{\pi}_{y0}$ )		...following sensitive behavior ( $\hat{\pi}_{y1}$ )	
	<i>Estimate</i>	<i>Std. Error</i>	<i>Estimate</i>	<i>Std. Error</i>
0	0.041	0.004	0.004	0.005
1	0.381	0.010	0.040	0.013
2	0.323	0.012	0.053	0.010
3	0.097	0.008	0.026	0.006
4	0.021	0.004	0.014	0.002
Total	0.863		0.137	
P-value	1			
<b>Panel B: No liar effects</b>				
	Information criterion			
	<i>BIC</i>	<i>AIC</i>	<i>BIC</i>	<i>AIC</i>
No boundary	9863.88	10050.28	9873.01	9885.44
Ceiling	9875.50	10173.80	9897.23	9894.14
Floor	9894.20	10176.84	9899.40	9915.55
Ceiling-Floor	9896.91	10300.36	9927.55	9924.26
Covariates	No	No	Yes	Yes

**Notes:** Panel A reports the estimated proportion of respondent types as described in each column title. The design effects test evaluates whether the population proportions are jointly non-negative. For each experiment, the Bonferroni-corrected  $P$ -value for the null of no design effects is reported. Panel B reports Schwarz’s (BIC) and Akaike’s (AIC) information criteria when the model is estimated without including boundaries (*No boundary*), including ceiling effects (*Ceiling*), including floor effects (*Floor*) and including both ceiling and floor effects (*Ceiling-Floor*). In this panel, the first two columns estimate the models without covariates, while the final two columns include the set of characteristics listed in Table 2.

<sup>29</sup> We also find similar results using a different set of covariates.

## A.4 Simple regression analysis

In the main text, we focus on the extreme bounds methodology to examine which variables are robustly correlated with tax evasion. This section reports a simpler regression analysis, which produces similar conclusions.

Table A-5 runs linear regressions for VAT evasion on the same set of variables explored in the text. Odd columns, labeled “bivariate”, show the resulting coefficient for regressions including only one covariate at a time (in addition to region fixed effects, which are always included). Even columns show the coefficient for a multivariate regression, which simultaneously includes all variables listed in the table. The reported significant correlations (and their magnitudes) fall in line with those that survive the sensitivity analysis with the extreme bounds methodology reported in the main text. Finally, in the main text we also explored the role of a few interactions between correlates of interest. In Table A-6 we show the results of including such interaction terms in regressions that include only region fixed effects and the relevant lower-order uninteracted terms (in the even, “bivariate” columns) as well as in regressions containing the full set of covariates Table and A-5. Again, there are few differences relative to the results using the extreme bounds methodology.

**Table A-5: Correlates of tax evasion**  
Simple regression analysis

	(1)	(2)		(3)	(4)
<i>Variables</i>	<b>Bivariate</b>	<b>Multivariate</b>	<i>Variables</i>	<b>Bivariate</b>	<b>Multivariate</b>
Wealth	-0.122*** (0.0219)	-0.0614** (0.0257)	Evangelical/Pentecostal	-0.0230* (0.0119)	-0.0331 (0.0292)
Use of violence	0.0748*** (0.0140)	0.0468*** (0.0148)	Fractionalization	-0.0368** (0.0169)	-0.0175 (0.0562)
Lands	-0.0443*** (0.00961)	-0.0373*** (0.00993)	Catholic	0.0190 (0.0125)	-0.0133 (0.0329)
Agree with bribery	0.0671*** (0.0137)	0.0455*** (0.0136)	Commerce sector	0.0218 (0.0162)	0.0239 (0.0181)
Guerrillas	0.0697*** (0.0135)	0.0601*** (0.0168)	Own welfare	-0.0138 (0.0130)	-0.0120 (0.0131)
Justice into own hands	0.0716*** (0.0133)	0.0335** (0.0153)	Age	-0.0144 (0.0128)	-0.0106 (0.0142)
Negative reciprocity	0.0686*** (0.0153)	0.0369** (0.0165)	Independent	0.0143 (0.0129)	0.00845 (0.0131)
Rural population	0.0900*** (0.0226)	0.0411 (0.0315)	Woman	0.0152 (0.0160)	0.0217 (0.0168)
Employment firms 2+	-0.0872*** (0.0230)	-0.0325 (0.0380)	Pop. density	-0.00926 (0.0164)	0.0412** (0.0200)
Education	-0.0456*** (0.0131)	-0.0233 (0.0157)	Polarization	0.0190 (0.0181)	-0.0580 (0.0479)
Other religion	-0.0275*** (0.0107)	-0.0329** (0.0162)	Neighbor cell phones	0.00841 (0.0155)	0.0159 (0.0158)
Win margin	0.0450** (0.0175)	0.0486 (0.0351)	Neighbor loans	-0.00666 (0.0144)	-0.00527 (0.0147)
State presence	-0.0401*** (0.0149)	-0.0277* (0.0153)	Homicide rate	0.00638 (0.0198)	-0.0349 (0.0216)
Shock	0.0282** (0.0122)	0.0223* (0.0120)	Gov. against inequality	-0.00593 (0.0151)	-0.00273 (0.0151)
HH expenses	-0.0411*** (0.0143)	-0.0147 (0.0148)	Paramilitaries	-0.00358 (0.0140)	-0.000216 (0.0153)
Popular vote	-0.0287** (0.0135)	-0.0226* (0.0136)	Positive reciprocity	0.00372 (0.0134)	0.00197 (0.0139)
Authorities violate law	0.0358*** (0.0134)	0.00518 (0.0142)	Government role	0.00319 (0.0145)	-0.00567 (0.0147)
Workers by firm	-0.0632*** (0.0185)	-0.0170 (0.0273)			

**Notes:** Ordinary least squares regressions. The dependent variable of interest is a dummy indicating whether the respondent normally accepts buying items without a receipt, to avoid paying VAT. Region fixed effects are always included, and standard errors are clustered at the community level. Odd columns, labeled “bivariate,” show the resulting coefficient for regressions including only one covariate at a time. Even columns show the coefficient for a multivariate regression, simultaneously including all variables listed in the table. For variable definitions, see Appendix Table A-1. \* is significant at the 10% level, \*\* is significant at the 5% level, \*\*\* is significant at the 1% level.

**Table A-6: Tax evasion: interaction terms**  
**Simple regression analysis**

	(1)	(2)
<i>Variables</i>	<b>Bivariate</b>	<b>Multivariate</b>
State presence*Positive rec.	-0.0221** (0.00991)	-0.0180* (0.0102)
State presence*Negative rec.	0.00753 (0.0127)	0.00570 (0.0130)
Popular vote*Positive rec.	0.00668 (0.0127)	0.00674 (0.0126)
Popular vote*Negative rec.	0.00503 (0.0140)	0.00260 (0.0136)

**Notes:** Ordinary least squares regressions. The dependent variable of interest is a dummy indicating whether the respondent normally accepts buying items without a receipt, to avoid paying VAT. Standard errors are clustered at the community level. Region fixed effects are always included, and standard errors are clustered at the community level. Column 1 reports the coefficient of a “bivariate regression” containing only the region fixed effects, lower-order uninteracted terms, and the interaction of interest as regressors. Column 2 presents the results of a multivariate regression in which all variables in Table A-5 are also included. For variable definitions, see Appendix Table A-1. \* is significant at the 10% level, \*\* is significant at the 5% level, \*\*\* is significant at the 1% level.